THE MOLECULAR ARCHITECTURE OF *MAMESTRA CONFIGURATA* PERITROPHIC MATRIX

A Thesis Submitted to the College of

Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

in the Department of Biology

University of Saskatchewan

Saskatoon

By

Umut Toprak

"Science is the most genuine guide in life."

Mustafa Kemal Atatürk

"Bugs are not going to inherit the earth. They already own it. It is time to make peace with the landlord."

Thomas Eisner

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from the University of Saskatchewan, I agree that the libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work, or in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or in part should be addressed to:

Head of the Department of Biology
University of Saskatchewan
Saskatoon, Saskatchewan, S7N 5E2, Canada

DEDICATION

This Ph.D thesis is dedicated to Pieter Lyonet, Edouard-Gérard Balbiani, Vincent Wigglesworth, Douglas Waterhouse, Robert Granados and all the researchers who allow me to build my story over their story. This thesis is also dedicated to all of the people working to find creative solutions to the challenges we face as inhabitants of planet earth.

ABSTRACT

The peritrophic matrix (PM) lines the insect midgut and is composed of chitin and protein. It is required for organization of digestion and for protection of epithelial cells from mechanical damage, pathogens, and toxins. The PM of *Mamestra configurata* (Lepidoptera: Noctuidae), bertha armyworm, a serious pest of cruciferous oilseed rape, was studied. The multilayered PM is delaminated from the anterior midgut epithelium during molting Phase II by periodic pulses and degraded during the molting Phase I stage. These events are controlled by chitin synthase-B, and chitinolytic enzymes, such as chitinase and β-*N*-acetylglucosaminidase. Eighty-two PM proteins were identified and classified as: i) peritrophins, ii) enzymes and iii) other proteins. Peritrophins were further classified as simple, binary, complex and repetitive according to their structural organization and phylogenetic analysis of peritrophin A domains. The expression of most genes encoding PM proteins was specific to the midgut and independent of larval feeding status, developmental stage, or PM formation.

This study includes the first report of chitin deacetylase (CDA) activity in the insect midgut suggesting that the PM may contain chitosan. Digestive enzymes, such as insect intestinal lipases (IILs) and serine proteases were also associated with the PM. The IIL genes differed in their expression during larval development; however, serine protease genes were expressed continuously and serine protease activity was present in the midgut of feeding and nonfeeding stages. *M. configurata* IIM4, a complex peritrophin, was susceptible to degradation by Mamestra configurata nucleopolyhedrovirus-A challenge, as the first evidence of IIM degradation by an alphabaculovirus enhancin. *M. configurata* IIM2, a binary peritrophin, was unaffected by baculoviral challenge and such resistance of an IIM has not been reported previously. The current study is also the first demonstration of silencing by RNA interference (RNAi) of any gene encoding a PM protein, in this case *M. configurata* CDA1 (McCDA1) and McPM1. In addition, both *in vitro* and *per os* feeding experiments revealed *McCDA1* silencing starting at 24 or 36 hours posttreatment, as one of the most successful demonstrations of RNAi in a lepidopteran.

ACKNOWLEDGMENTS

I owe my deepest gratitude to my supervisors, Drs. Martin Erlandson and Cedric Gillott, and to Dr. Dwayne Hegedus, Cognate Member of my Advisory Committee, for the ongoing encouragement, guidance, and patience they have provided for me from the initial to the final stages of my Ph.D studies. Their support was truly inspirational.

I would like to thank my committee members Drs. Peta Bonham-Smith and Neil Chilton for their guidance and constructive feedback and Dr. Michel Cusson for serving as my external examiner.

A special thanks to Doug Baldwin, whose assistance and friendship helped me tremendously both in the lab and in my life in Saskatoon. I also thank Cathy Coutu, Stephanie Harris, Alison Paton, Keith Moore and Diney Bekkaoui for their day to day support for making things go smoothly as much as possible. I thank Drs. Guosheng Liu and Branimir Gjetvaj for their assistance on microscopy.

This thesis would not have been possible without the emotional support of my sister Sinem Bertling. Her mental guidance was truly inspirational to me. I also would like to acknowledge the continuous support of my mother Hayrünnisa Toprak and father Baki Toprak. I would like to thank my girlfriend Fehime Utkan for believing in me and for having my back at difficult times. I am also grateful to Dr. Oktay Gürkan for his continuous support and mentorship from Turkey. Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Ac.....*Autographa californica*

ACT.....Actin

Ag.....*Anticarsia gemmatalis*

Ai.....*Agrotis ipsilon*

AlkPPc .....Alkaline phosphatase domain

ALP .....Alkaline phosphatase

AMY .....α-Amylase

APN.....Aminopeptidase-N

AST .....Astacin

AWGA .....Alexafluor 488 conjugated wheat germ agglutinin

BAC .....Bacterial artificial chromosome

Bm.....*Bombyx mori*

Bn .....*Brassica napus*

BsPgDA .....*Bacillus subtilis* chitin deacetylase

Bt.....*Bos taurus*

BV .....Budded virus

CBD .....Chitin binding domain

CBD3P .....Chitin binding domain 3 protein

CBP .....Chitin binding pocket

CDA .....Chitin deacetylase

Cf.....*Choristoneura fumiferana*

CHAPS.....Cholamidopropyl dimethylammonio propanesulfonate

CHI.....Chitinase

CHS.....Chitin synthase

ClCDA .....*Colletotrichum lindemuthianum* chitin deacetylase

CLECT .....C type lectin

CPA.....Carboxypeptidase A

CPB .....Carboxypeptidase B

CRD .....Carbohydrate recognition domain

CtAXE.....*Clostridium thermocellum* chitin deacetylase

CTLP.....Chymotrypsin like protease
DEPC .....Diethyl pyrocarbonate
Dm.....*Drosophila melanogaster*
Dp.....*Drosophila pseudo-obscura*
dsRNA.....double-stranded RNA
dsRNase .....double-stranded RNase
Ep .....*Epiphyas postvittana*
EST .....Expressed sequence tag
FB.....Fat body
FB28.....Fluorescent Brightener 28
FG .....Foregut
Fp .....Forward primer
GAG.....Glycosaminoglycan
GlcNAc .....β-(1, 4)-*N*-Acetyl-D-glucosamine
$GlcNH_2$.....D-Glucosamine
GmLe1 .....Soybean lectin
Gor .....Glutathione reductase
GPI.....Glycosylphosphatidylinositol
GSP .....Gene specific primer
GV.....Granulovirus
Ha.....*Helicoverpa armigera*
Hc.....*Hyphantria cunea*
HG.....Hindgut
HMG176 .....*Helicoverpa* midgut 176
hpi .....Hours post inoculation
hpt .....Hours post treatment
IIL .....Insect intestinal lipase
IIM .....Insect intestinal mucin
IN .....Integument
LC .....Larval carcass
LC-MS/MS .....Liquid chromatography tandem-mass spectrometry

Ld .....*Lymantria dispar*

LD_{50}.....Lethal dose fifty

Lo .....*Lacanobia oleracea*

Lsti .....*Loxostege sticticalis*

Maco .....*Mamestra configurata*

Mb .....*Mamestra brassicae*

Mc .....*Mamestra configurata*

MG176 .....Midgut 176

MNPV .....Multiple nucleopolyhedrovirus

Mise.....*Mythimna separata*

MG .....Midgut

Mow .....Mowse

MT.....Malpighian tubules

Ms .....*Manduca sexta*

NAG.....β-*N*-acetylglucosaminidase

NAP.....Napin

NCBI .....National Center for Biotechnology Information

NPV.....Nucleopolyhedrovirus

OB .....Occlusion body

Obst .....Obstructor

ODV .....Occlusion derived virion

Of .....*Ostrinia furnacalis*

ORF.....Open reading frame

PAD.....Peritrophin A domain

PAN.....Pantetheinase

PBD.....Peritrophin B domain

PBS .....Phosphate buffered saline

PCD.....Peritrophin C domain

Pep.....Peptide

PM.....Peritrophic Matrix

POL .....Polycalin

PPAD .....Protein with Peritrophin-A domain
Pssm .....Position specific scoring matrix
qPCR .....Quantitative real-time PCR
RACE .....Rapid Amplification of cDNA Ends
rCDA .....Recombinant chitin deacetylase
REPAT .....Response to pathogen
RhiNODB .....*Sinorhizobium meliloti* chitin deacetylase
rMcPPAD .....Recombinant McPPAD
RNAi .....RNA interference
Rp .....Reverse primer
RT-PCR .....Reverse transcription-PCR
SAHS .....Sterile artificial hemolymph solution
Se .....*Spodoptera exigua*
SEM .....Scanning electron microscopy
Seq cov .....Sequence coverage
Serpin .....Serine protease inhibitor
Sf .....*Spodoptera frugiperda*
siRNA .....Small-interfering RNA
Sl .....*Spodoptera litura*
SlAXE .....*Streptomyces lividans* chitin deacetylase
SP .....Serine protease
SpPGDA .....*Streptococcus pneumoniae* chitin deacetylase
TBP .....Tributyl phosphine
Tc .....*Tribolium castaneum*
Tn .....*Trichoplusia ni*
TR .....Tracheae
TrxB .....Thioredoxin reductase
TUB .....Tubulin
UDP-GlcNAc .....Uridine diphosphate-*N*-acetylglucosamine
UTR .....Untranslated region
Xn .....*Xestia c-nigrum*

wt.....Wild type

β1, 3GLU β-1, 3-glucanase

1D.....One dimensional

2D .....Two dimensional

1. GENERAL INTRODUCTION

1.1 The Insect Gut

In nature, insects exploit a wide diversity of food sources. Although these differences in feeding habits lead to significant variability in gut structure and function, the insect gut is in almost all species a continuous tube throughout the body and primarily divided into three regions; the foregut (stomodeum), the midgut (mesenteron) and the hindgut (proctodeum) (Gillott, 2005) (Figure 1.1). The foregut's main roles are ingestion, storage, grinding and transportation of food to the midgut. The midgut is the principal site for secretion of digestive enzymes and for digestion and absorption of digested products. The hindgut regulates the absorption of water and salts from waste products in the alimentary canal.

Insects protect themselves against various harmful biotic and abiotic factors in the environment by their cuticular integument (Wang and Granados, 2001). The foregut and the hindgut are also lined with a protective cuticular layer called the intima; however, the midgut lacks this layer. Instead, epithelial cells at the foregut midgut junction or along the entire midgut produce a semipermeable, sheath-like, acellular structure called the peritrophic matrix (PM) which surrounds the food bolus. The PM separates the midgut into two regions; the endoperitrophic space which contains the food bolus, and the ectoperitrophic space which is the region between the PM and the midgut epithelium (Figure 1.1). The PM may also line the hindgut and is later eliminated with the fecal pellets (Brandt et al., 1978).

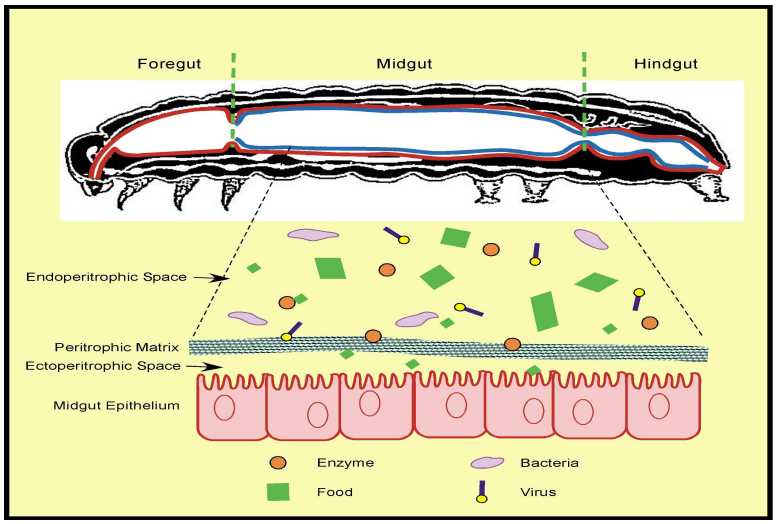


Figure 1.1 The alimentary canal of a lepidopteran larva. (Source: Sarauer, 2002).

1.2 The Peritrophic Matrix

The presence of a PM in insects was first noted by Lyonet (1762; cited in Lehane, 1997) over two centuries ago, and was first referred to as the "peritrophic membrane" by Balbiani (1890; cited in Lehane, 1997). However, the term membrane may be confused with the bilipid membrane surrounding a cell, and the term "peritrophic matrix", which was first proposed by Ramos et al. (1994), more appropriately describes this structure (Jacobs-Lorena and Oo, 1996). In several articles, the term "peritrophic envelope" is also used. This term was proposed by Peters (1992) due to the multilayer structure detected in the PMs of several insect orders; however, as either a membrane or a matrix could be composed of multiple layers, this change in the terminology was considered unnecessary (Terra, 2001).

1.2.1 Structure

PM is mainly composed of chitin and a distinct group of proteins. The relative amounts of these components vary among species and life stages as well as among the maturation states of the PM within one life stage (Tellam, 1996).

1.2.1.1 Chitin

Chitin, a polymer of β-(1, 4)-*N*-acetyl-D-glucosamine (GlcNAc), constitutes between 3 and 13% of the PM (Peters, 1992; Tellam and Eisemann, 2000). Chitin synthase-B (CHS-B) catalyzes the synthesis of chitin in the PM by the sequential addition of GlcNAc units to the elongating end of the growing chitin chains which then organize into microfibrils (Arakane et al., 2005; Bolognesi et al., 2005). Chitin microfibrils have uniform diameters in the range of 2-10 nm and exceed 500 nm in length (Lehane, 1997). These are often arranged in parallel groups of 10 or more to form bundles (Figure 1.2a-1.2b) (Lehane, 1997; Jang et al., 2004). This organization makes the PM appear as a mesh like or lattice structure (Figure 1.2c). Chitin is generally formed at the tips of midgut microvilli and the internal spaces within the orthogonal lattice are slightly larger than the microvilli, suggesting that microvilli serve as a template for the macro-organization of the PM (Harper and Hopkins, 1997).

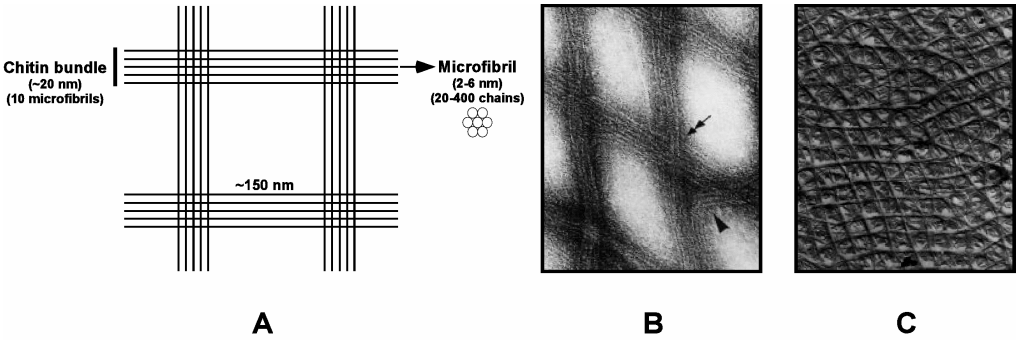


Figure 1.2 Organization of chitin microfibrils that form the peritrophic matrix. A: Illustration of chitin microfibrils and bundles. B: Negative staining of chitin microfibrils within microfibrillar bundles of the PM of *Forficula auricularia* (magnification: 270, 000X). Regions of felting (double arrowhead) and single microfibrils sharply changing course (single arrowhead) are shown. C: Negative staining of microfibrillar bundles within the PM of *Tipula* sp. larva (Magnification: 22,000X). (Source: Peters, 1992).

1.2.1.2 Proteins

Proteins account for a significant portion (21-55%) of the total mass of the PM (Peters, 1992; Tellam, 1996). The first structural PM protein characterized, Peritrophin-44, is a small chitin binding glycoprotein from the larvae of Australian sheep blowfly, *Lucilia cuprina* (Elvin et al., 1996). Subsequently, expressed sequence tag (EST) analysis of midgut epithelial tissues (Shi et al., 2004; Simpson et al., 2007; Venancio et al., 2009) and proteomic analysis of the PM (Campbell et al., 2008; Dinglasan et al., 2009) and midgut lumen components (Pauchet et al., 2008) have substantially increased the number of characterized PM proteins. Most of these studies focused on the PM proteins from dipterans and lepidopterans and less commonly from coleopterans and hymenopterans.

Proteins bound to the PM have been divided into four classes based on the ease of their release from the PM (Tellam et al., 1999). Class I proteins are loosely bound to the PM and can be released by very mild treatments, such as rinsing with low or high ionic strength buffers. Class II proteins can be removed from the PM by treatment with mild detergents (e.g., 2% Triton X-100 or 2% Zwittergent 3-14) that disrupt weak protein-protein, protein-oligosaccharide or protein-chitin interactions. In contrast, Class III proteins are extractible only with the use of strong denaturants such as urea or guanidine hydrochloride under nonreducing conditions. Class III proteins are termed “peritrophins” to emphasize their strong association within the PM. Proteins remaining in the nonextractible residue after these treatments are referred to as Class IV and are likely associated through some form of covalent interaction.

Now that many PM proteins have been characterized, they are more appropriately classified according to their function as either structural proteins or enzymes. Structural proteins, referred to as peritrophins, contribute to the major physical characteristics of the PM including strength, elasticity and porosity and are responsible for many PM functions. Peritrophins bind to chitin (Schorderet et al., 1998; Wang et al., 2004a), a characteristic attributed to the presence of one or more chitin binding domains (CBDs) in which aromatic amino acid R-groups form hydrogen bonds with the pyranose ring of GlcNAc (Kraulis et al., 1989; Xu et al., 1995). Peritrophin CBDs are classified as either Type A (Peritrophin A Domain: PAD, also known as Type 2), Type B (PBD) or Type C (PCD) and have a register of 6, 8 or 10 cysteines, respectively, that form 3-5 intradomain disulphide bonds (Tellam et al., 1999). Peritrophins contain at least one CBD but may also have a mucin domain (MD) in which case they are often referred to as insect intestinal mucins (IIMs) (Wang and Granados, 1997b). MDs contain regions that are rich in threonines and serines and show a high potential for *O*-linked glycosylation (Devine and McKenzie, 1992). The carbohydrate moieties associated with the MDs may account for 50-90% of the protein by weight (Perez-Vilar and Hill, 1999). Prolines are also spread among the threonines and serines which cause the amino acid chain to condense into a coiled structure with the carbohydrate moieties projecting outward (Van Klinken et al., 1995). Many dipteran peritrophins have a simple CBD organization with only a few domains (Shen and Jacobs-Lorena, 1998; Devenport et al., 2005; Shao et al., 2005) whereas lepidopteran peritrophins generally have multiple CBDs and MDs (Wang et al., 2004a; Shi et al., 2004).

Several types of enzymes have been also reported to be associated with the PM. These include digestive enzymes (e.g., serine proteases, lipases, exopeptidases and amylases), and chitin modifying enzymes [e.g., chitin deacetylases (CDAs)] (Terra et al., 1979; Campbell et al., 2008). Indeed, the first proteins reported from the PM were serine proteases and exopeptidases (Terra et al., 1979; Walker et al., 1980), while the association of CDAs with the PM was reported only recently (Guo et al., 2005). While the role of digestive enzymes within the PM is obvious, that of CDAs is not. As the name suggests, CDAs catalyze the release of acetyl groups from chitin, forming deacetylated chitin or chitosan (Tsigos et al., 2000). A recombinant nucleopolyhedrovirus (NPV) expressing a CDA associated with the *Helicoverpa armigera* PM was found to be more virulent and to increase PM permeability, suggesting deacetylation of chitin leads to changes in PM structure (Jakubowska et al., 2010).

Several other proteins, such as lipocalins, serpins, alkaline phosphatases and β-1, 3-glucanases, have been also reported to be associated with insect PMs (Campbell et al., 2008).

1.2.2 Formation of PM

Two types of PM have been described based on the site of chitin secretion (Peters, 1992). Type I PM is formed along the entire length of the midgut epithelium, whereas Type II PM is formed by either a special ring of cells or the cardia near the foregut-midgut boundary (Wigglesworth, 1972). Type I PM can be continuously present as in most lepidopteran larvae (with the possible exception of early pharate larvae) or formed in response to feeding, as in most haematophagous female dipteran adults (Waterhouse, 1953b; Wigglesworth, 1972). The structure of the Type I PM may vary among different orders or life stages. In haematophagous dipterans, PM forms as a bag like structure containing the ingested meal and a single or multiple PMs may be produced depending on the frequency of blood feeding (Waterhouse, 1953a). The concentric overlapping PMs sometimes observed in other insects, such as locusts (Baines, 1978), may also arise from periodic pulses of PM synthesis. Type I PM is also present in beetles (Coleoptera), dragonflies (Odonata), mayflies (Ephemeroptera), and stick and leaf insects (Phasmida) (Waterhouse, 1957; Terra, 2001).

Type II PM is produced continuously in dipteran larvae (Wigglesworth, 1930) and a few lepidopteran larvae or adults (Waterhouse, 1953b; Ryerse et al., 1992). Type II PM is secreted initially as a fluid, which is then moulded to form the PM sleeve (Wigglesworth, 1929; 1930). It is noteworthy that some insects may have both types of PM formation in either the same or different developmental stages. For example, in cockroaches although an initial PM is formed by the midgut caeca, multiple PMs eventually arise along the entire length of the midgut epithelium indicative of a combination of Type I and Type II PM formation styles (Lee, 1968; Zhuzhikov, 2003). Similarly, ant (*Myrmica* sp.) (Weir, 1957) and honey bee (Snodgrass, 1956) larvae were reported to produce both Type I and II PMs at different developmental stages.

1.2.3 Peritrophic Matrix Functions

As a barrier between the food and the midgut epithelium, the PM has a number of functions; these are summarized below:

1.2.3.1 Mechanical protection and serving as a potential barrier to pathogens

The primary role of the PM is to serve as a mechanical barrier to prevent damage from abrasive food particles (Waterhouse, 1953b). Evidence for this was obtained from a *Bombyx mori* mutant that lacked a PM in the larval stage. The midgut epithelium of feeding mutant larvae was highly abraded and membranous bodies were released from its surface into the bolus (Sudha and Muthu, 1988). In *Ostrinia nubilalis* larvae, disruption of PM formation by feeding on wheat germ agglutinin led to the disintegration of microvilli as food particles penetrated into this layer (Harper et al., 1998; Hopkins and Harper, 2001). Feeding was reduced in these larvae, suggesting that they are attempting to avoid injury to the midgut epithelium (Hopkins and Harper, 2001).

Lehane (1997) reported that a major role of the PM is to serve as a barrier against pathogens ingested with the food. Thus, insects that naturally feed exclusively on a largely sterile diet tend to lack a PM; however, those that feed on diets heavily contaminated with microorganisms often have a PM (Lehane, 1997). Interestingly, butterfly species that feed on highly contaminated foods have a Type II PM, whereas those that feed on less contaminated foods produce no PM or a Type I PM (Waterhouse, 1953b), suggesting Type II PMs may be a more effective barrier to pathogens than Type I PMs (Lehane, 1997).

Bacillus thuringiensis produces endotoxins that disrupt epithelial cells and facilitate infection; however, the toxins were found to be entrapped by the glycan moieties on PM proteins (Hayakawa et al., 2004), showing that the PM can indirectly diminish the efficacy of such pathogens. Levy et al. (2007) examined the morphology of the PM from *Anticarsia gemmatalis* strains that were susceptible to infection with Anticarsia gemmatalis multiple nucleopolyhedrovirus (AgMNPV). They found that the PM in resistant larvae was thicker, had higher chitin content and was composed of several layers of fibrous/vesicular materials compared to the PM from susceptible larvae, indicating that the structure and composition of the PM is a factor in resistance to AgMNPV infection. In another study, Plymale et al. (2008) reported that the thickness of the PM was dependent on diet and this contributed to the susceptibility of *Heliothis virescens* larvae to infection by Autographa californica multiple nucleopolyhedrovirus (AcMNPV). There are circumstances where the PM may not be an effective barrier to infection; for example, specific baculoviruses (Derksen and Granados, 1988;

Wang and Granados, 1997a) and entomopoxviruses (Mitsuhashi and Miyamoto, 2003; Mitsuhashi et al., 2007) degrade PM structural proteins or disrupt PM architecture.

1.2.3.2 Permeability

The porous structure of the PM allows movement of digestive enzymes from midgut epithelial cells through the PM into the endoperitrophic space and the retrograde movement of digested food and smaller molecules from lumen to the ectoperitrophic space (Peters, 1992). Pore size alone results in a certain degree of selectivity; thus, larger molecules, bacteria, viruses and parasites are retained in the gut lumen (Tellam, 1996).

PM permeability depends on other physical and chemical characteristics. The PM has a fixed charge density determined largely by its glycosaminoglycan (GAG) content, which is contributed by polysaccharides on IIMs. The packing of the polysaccharides is affected by both pH and the ionic environment that in turn affect permeability by altering ionic interactions between the solute and the PM (Lehane, 1997). For example, the presence of calcium ions increased the permeability of the PM in the tsetse fly, *Glossina morsitans morsitans* (Miller and Lehane, 1993). Alteration of the midgut pH could also affect the density of anionic charges associated with the GAGs thereby altering the permeability of the PM. This could be particularly important in insect species where different regions of the gut have markedly different pH values (Tellam, 1996). For instance, permeability of the *Manduca sexta* larval PM was lower at the posterior end (pH 7.1) versus the anterior end (pH 7.7) and also varied with the developmental stage (Dow and O'Donnell; 1990; Spence and Kawata, 1993). Details of the molecular mechanisms that determine the permeability of the PM are not clear and several discrepancies have been reported. For example, the PM of *Anopheles stephensi* larvae was permeable to fluorescent labelled dextrans (2.4-17.2 kDa); however, the PM of *Aedes aegypti* larvae was impermeable even to the smallest dextrans (2.4 kDa) (Peters and Wiese, 1986). Jakubowska et al. (2010) demonstrated an increase in PM permeability after inoculation of *H. armigera* larvae with a recombinant baculovirus expressing CDA, indicating that deacetylation of chitin in the PM affects PM permeability.

1.2.3.3 Compartmentalization of digestive processes

As mentioned above, the PM effectively divides the midgut lumen into an endoperitrophic space and an ectoperitrophic space. The initial stages of digestion occur within the endoperitrophic space where large macromolecules (e.g., proteins, complex carbohydrates, and lipids) are reduced to smaller molecules, such as peptides, sugars, and fatty acids. The final stages of digestion and nutrient uptake occur within the ectoperitrophic space where small polymers are reduced to their molecular components and taken up by epithelial cells (Terra, 2001). This compartmentalization allows smaller enzymes, such as amylase and trypsin, to pass through the PM into the endoperitrophic space to initiate digestion, while larger and epithelium-bound enzymes, such as carboxypeptidases, aminopeptidases and some glycosidases, are retained in the ectoperitrophic space (Terra and Ferreira, 1981). A counter-current water flow in the ectoperitrophic space was postulated to shuttle digestive enzymes toward the anterior midgut where they again pass across the PM into the endoperitrophic space (Ferreira et al., 1981). Thus, countercurrent flow would allow digestive enzymes to be recycled and digestion products to be absorbed by columnar epithelial cells along the entire length of the midgut (Caldeira et al., 2007).

1.2.3.4 Protection from ingested toxic compounds

The PM may protect some phytophagous insects from the toxic effects of ingested phenolic compounds commonly found in many plants (Barbehenn and Martin, 1994). This can occur through binding of toxic material, ultrafiltration, polyanion exclusion and antioxidant activity (Barbehenn, 2001). Recently, *B. thuringiensis* Cry toxins were demonstrated to bind to the PM of toxin-resistant lepidopterans (*Agrotis ipsilon*, *Mamestra brassicae* and *Spodoptera exigua*) *via* interaction with IIMs which prevented them from reaching their target sites on midgut epithelium (Rees et al., 2009). In *Schistocerca gregaria*, up to 30% of the potentially toxic dietary tannins were bound to PM and shed in the feces (Bernays and Chamberlain, 1980). IIMs may be involved in detoxification as the IIM AeIMUC1 was produced by the midgut epithelium in response to heavy metal exposure in *Aedes aegyptii* larvae (Rayms-Keller et al., 2000). In adult female mosquitoes, the iron containing heme binds to AeIMUC1 in the PM during the process of hemoglobin degradation, thus limiting accumulation of toxic heme groups (Pascoa et al., 2002; Devenport et al., 2006). Adsorption of iron to the PM was also shown in the

caterpillar, *Malacosoma disstria*, when the larvae were fed increasing amounts of dietary iron (Barbehenn and Stannard, 2004). PMs may also serve as a molecular sieve and physically exclude molecules through a process known as ultrafiltration (Barbehenn, 2001). Negatively charged molecules are repelled by the similarly charged PM, referred to as polyanion exclusion (Barbehenn, 2001). In *Helicoverpa zea*, the PM was demonstrated to scavenge hydroxyl radicals *in vitro* and reduce the formation of hydroperoxides in midgut tissues exposed to a hydroxyl radical generator, suggesting the PM may have antioxidant functions (Summers and Felton, 1996). When PM formation was inhibited by calcofluor in *M. disstria* and *Orgyia leucostigma*, protein carbonyls, the most widely used marker of oxidative modification of proteins, increased by two- to three-fold in the midgut epithelium confirming the radical scavenging antioxidant feature of the PM (Barbehenn and Stannard, 2004). As such, the PM was considered to be a "sacrificial antioxidant".

1.2.4 The Peritrophic Matrix as a Target for Novel Methods of Insect Control

The PM is an attractive target for pest management strategies due to its intimate involvement in digestive processes and role as a first line of defence against pathogens. Compounds that interact with the PM have the potential to affect its physiochemical and biological properties. A number of strategies that target the PM as a form of insect control are outlined below.

1.2.4.1 Compounds that block or inhibit the passage of nutrients resulting in retardation of larval growth

A variety of compounds may block PM pores thus compromising nutrient uptake and digestive efficiency (Eisemann et al., 1994; Casu et al., 1997). Lectins, small carbohydrate binding proteins, have been reported to interfere with insect growth and development and have been suggested for use as selective tools for insect control (Czapla and Lang, 1990; Murdock et al., 1990). For example, ingestion of wheat germ agglutinin decreased PM permeability, reduced growth rate, and caused mortality of *L. cuprina* larvae (Eisemann et al., 1994). Interaction of lectins with the PM may block nutrient passage in pores and decrease digestive efficiency by preventing the passage of enzymes into the lumen and nutrients to the ectoperitrophic space. *O. nubilalis* larvae fed wheat germ agglutinin for 24 h hypersecreted a multilayered abnormal PM in

the anterior midgut and disintegration of microvilli was observed (Harper et al., 1998). Hypersecretion of PM in the anterior region of the midgut lumen and reduced intake of diet may be a defensive reaction to injury to the midgut cells (Hopkins and Harper, 2001) and therefore such insects may be less affected by lectins.

PM proteins may be candidate antigens for the development of antibodies that could block PM pores. Casu et al. (1997) reported that an antibody developed against *L. cuprina* Peritrophin-95 within the immune system of the host (sheep) inhibited the growth of *L. cuprina* larvae. They suggested that ingested antibodies to Peritrophin-95 bound to the *L. cuprina* PM and interfered with the movement of nutrients to the midgut epithelium. Mosquitoes feeding on hosts that had been immunized with mosquito gut material showed reduced PM formation (Ramasamy et al., 1996) and fecundity (Sutherland and Ewen, 1974) and increased mortality (Alger and Cabrera, 1972).

1.2.4.2 Compounds interfering with PM formation

Several studies reported that optical brighteners interfere with chitin assembly (Herth, 1980; Bartnicki-Garcia et al., 1994; Garcia-Zapien et al., 1999). This finding led to studies on the use of these compounds to affect the efficacy of entomopathogens that infect insects *per os*. Most of these studies revealed that optical brighteners enhance the efficacy of baculoviruses, though little is known about the exact mechanism by which this occurs. Wang and Granados (2000) reported that calcofluor enhanced baculovirus infection by interfering with PM formation. Optical brighteners were shown to bind to newly synthesized chitin resulting in inhibition of protein binding to chitin in the PM (Okuno et al., 2003) and disruption of PM integrity (Shapiro and Dougherty, 1994; Wang and Granados, 2000). Zhu et al. (2007) showed that this disruption is due to the dissociation of proteins from the PM, which increased the susceptibility of PM proteins to degradation by digestive proteases. In other cases, a PM was not detectable when larvae were fed with calcofluor (Rees et al., 2009).

In another study, Spodoptera litura multiple nucleopolyhedrovirus virulence was found to be enhanced when *Spodoptera litura* larvae were fed with the chitin synthesis inhibitor, chlorfluazuron. This caused the PM to rupture which presumably facilitated the passage of more virions through the PM (Guo et al., 2007a).

1.2.4.3 Hydrolytic enzymes

Insect parasites or pathogens and plants can produce hydrolytic enzymes that disrupt the PM. For example, Derksen and Granados (1988) reported that three major glycoproteins in the PM of *Trichoplusia ni* larvae were degraded after challenge with Trichoplusia ni granulovirus (TnGV). Wang and Granados (1997a) demonstrated *in vivo* and *in vitro* that degradation of fully processed forms of TnIIMs was due to the activity of a specific metalloprotease, enhancin, encoded by some baculovirus genomes. In addition, *T. ni* and *P. unipuncta* PMs treated with TnGV enhancin were found to be significantly more permeable to fluorescent labelled AcMNPV than nontreated PMs (Peng et al., 1999). This is likely due to the physical disruption of the PM (Guo et al., 2007b; Hoover et al., 2010). Enhancin was also demonstrated to enhance baculoviral infections (Gallo et al., 1991; Wang and Granados, 1998; Del Rincon-Castro and Ibarra, 2005) and toxicity of *B. thuringiensis* (Granados et al., 2001). Li et al. (2003) reported that a recombinant AcMNPV containing a gene encoding Mamestra configurata nucleopolyhedrovirus (MacoNPV-A) enhancin had a significantly lower LD_{50} value for 2nd instar *T. ni* larvae, providing direct evidence of enhancement due to enhancin. Degradation of the *Mamestra configurata* IIM, McIIM1, was reported during MacoNPV-A *per os* infections; interestingly, only the unprocessed or partially processed forms of McIIM1 were degraded, while the fully glycosylated form was unaffected (Shi et al., 2004). This is in contrast to findings for TnGV in which the fully glycosylated form of TnIIM was degraded. Scanning electronic microscopy (SEM) showed that Xestia c-nigrum granulovirus enhancin disrupted the PM of *S. litura* larvae potentially facilitating the passage of virions through the PM (Guo et al., 2007b).

In haematophagous insects, many insect-vectored mammalian parasites must cross the PM to invade the midgut epithelium. This process is often facilitated by chitinases that degrade the PM (Huber et al., 1991), for example, those secreted by *Leishmania* and *Plasmodium* species (Shao et al., 2001; Abraham and Jacobs-Lorena, 2004). Chitinases have the potential to be used as tools to disrupt the PM and studies have shown that they have caused mortality or reduced the survival and growth of various insects (Wang et al., 1996; Fitches et al., 2004; Lertcanawanichakul et al., 2004). In addition, chitinases were also demonstrated to enhance the efficacy of baculoviruses (Shapiro et al., 1987). Other hydrolytic enzymes produced by plants, such as cysteine proteases, also disrupt the PM. Pechan et al. (2002) reported maize lines resistant to *H. zea* produced a cysteine protease that targets the PM. SEM indicated that the PM

of larvae reared on the leaves of resistant lines was severely damaged with cracks and holes appearing on the luminal side of the PM (Mohan et al., 2006). Konno et al. (2004) also reported that papain, a cysteine protease found in latex, is crucial in the defense of the papaya tree against lepidopteran larvae.

1.2.4.4 RNA interference (RNAi)

RNAi is a method that is used to deplete specific mRNAs by using a double-stranded (ds) RNA molecule. dsRNA can survive the biochemically hostile environment in the insect gut and is taken up by the midgut epithelial cells, leading to the elimination of transcripts from the target gene and a reduction of protein levels (Turner et al., 2006). The dsRNA also can be generated *in vitro* or *in planta* (Baum et al., 2007). Transgenic plants expressing a dsRNA against *CYP6AE14*, which encodes a P450 monooxygenase that detoxifies plant defence compounds (Li et al., 2002b) or insecticides (Waters et al., 1992), was shown to retard larval growth and development of *H. armigera* dramatically in the presence of gossypol (Mao et al., 2007). Genes encoding major PM proteins or enzymes involved in chitin regulation could also be targeted by this approach to disrupt PM formation. For example, the transcript level of a gene encoding a midgut chitinase was reduced (>60%) after feeding dsRNA to *O. nubilalis* larvae (Khajuria et al., 2010); however, the technique has been more successful in coleopterans and dipterans. In *Tribolium castaneum*, injection of dsRNA specific for *CHS-B* in penultimate instar larvae reduced larval size and chitin content in the PM and these larvae died (Arakane et al., 2005; 2008). In *Ae. aegypti* adults, PM formation was inhibited when *CHS-B* transcript levels were reduced using RNAi (Kato et al., 2006).

1.3 Objectives of the Study

As reviewed above, the PM has essential functions associated with digestive processes and protects epithelial cells as a first line of defence against pathogens and toxins. The overall goal of this research was to elucidate the formation and structure of the lepidopteran larval PM, and the contribution of some PM components to PM functions. These functions also make the PM an attractive target for development of novel *per os* insect control strategies. As such, the study also aims to determine the PM components that have potential as targets within an insect control strategy. The specific objectives of this research are:

1. To gain a better understanding of the morphology, histology and formation of the PM throughout larval development. This included characterization of the enzymes responsible for chitin synthesis and degradation (Chapter 3).
2. To survey proteins associated with the PM (Chapter 4).
3. To characterize select PM proteins in detail (Chapter 5).
4. To expand the understanding of proteins with Peritrophin-A Domains (PPADs), of which peritrophins are members (Chapter 6).
5. To examine the interaction between IIMs and baculovirus and the baculoviral metalloprotease, enhancin (Chapter 7).
6. To evaluate RNAi as a means to inhibit the expression of specific midgut genes encoding PM components using both *in vivo* and *in vitro* systems (Chapter 8).
7. To develop a model of the lepidopteran PM (Chapter 9).

2. GENERAL MATERIAL AND METHODS

This chapter presents the general material and common methodologies used in the entire study. Specific material and methods are presented in the appropriate chapters.

2.1 Research Material

2.1.1 Mamestra configurata

The insect model chosen for this study was *M. configurata*, for which there is already some information on its digestive biochemistry (Hegedus et al., 2003; Erlandson et al., 2010). This species was also used to develop the current model of the lepidopteran Type I PM (Shi et al., 2004).

M. configurata, commonly known as the bertha armyworm (BAW), is a noctuid moth found across western North America. It is a serious but sporadic pest of cruciferous oilseed rapes (*Brassica napus* and *B. rapa*) and several other crops (Turnock, 1985; Mason et al., 1998). The current control strategy for *M. configurata* consists of aerial application of chemical insecticides. In the last outbreak (in 1994 and 1995), 670,000 ha of canola were sprayed at a cost of $50 million and yet incremental yield losses of $60 million occurred in Canada (Mason et al., 1998).

Larvae were reared at 21±1 ºC under a 16 h light/8 h dark photoperiod and fed *ad libitum* on artificial diet (Bucher and Bracken, 1976), artificial diet supplemented with 0.5% (w/v) soybean Kunitz-type trypsin inhibitor (Sigma, St Louis, MO, USA) or *Brassica napus* (AC Excel). The genetic diversity of the colony was maintained by annually mating colony insects with moths derived from field collected pupae.

2.2 Midgut cDNA Library Construction and Expressed Sequence Tag Generation

Collection of larval midgut tissues and cDNA library construction were as described in Chamankhah et al. (2003). In total, 5797, 5866 and 4376 expressed sequence tags (ESTs) were generated from the midgut cDNA libraries from larvae feeding on diet, diet with soybean trypsin inhibitor and *B. napus* leaves, respectively. The data were warehoused using the software package APED (http://aped.sourceforge.net) (Links et al., 2007). DNA sequences were evaluated for quality and trimmed of vector sequences prior to assembly using Lucy (Chou and Holmes, 2001). Assembly of the data was performed using TGICL with a Cap3 (Huang and Madan, 1999)

parameter of –p98 (Pertea et al., 2003). The combined ESTs assembled into 1646 contigs and 3280 singleton sequences (Erlandson et al., 2010).

2.3 Rapid Amplification of cDNA Ends (RACE)

RACE reactions were performed to generate full length cDNAs using the 5' RACE System or GeneRacer™ Advanced RACE kit (Invitrogen, Carlsbad, CA, USA). In 5' RACE analysis, the cDNA templates were generated by reverse transcription using SuperScript™ II (5' RACE System kit) or SuperScript™ III (GeneRacer™ Advanced RACE Kit) with gene specific primers. The 5' ends of the templates were amplified from the cDNA with the appropriate forward primers (provided by the kits) and gene specific reverse primers using Taq DNA polymerase (5' RACE System) or Platinum® Taq high fidelity DNA polymerase (GeneRacer™ Advanced RACE). The PCR conditions used in the 5' RACE System were 94°C for 1.5 min, followed by 35 cycles of 94°C for 45 sec, 55°C for 45 sec, 72°C for 2 min and a final extension at 72°C for 6 min. The PCR conditions used in the GeneRacer™ Advanced RACE system were 94°C for 2 min, followed by 30 cycles of 94°C for 30 sec, 64°C for 30 sec, 68°C for 90 sec and a final extension at 68°C for 10 min. The PCR products were cloned into pGEM-T Easy (Promega, Madison, WI, USA) or pCR®4-TOPO® (Invitrogen) vectors in *Escherichia coli* DH10B or TOP10, respectively. 3' RACE was conducted using a forward gene specific primer and the GeneRacer™ oligo dT reverse primer with the same PCR conditions used for the GeneRacer™ Advanced RACE system.

2.4 RNA Extraction and Analysis of Gene Expression

Gene expression was examined using northern blot and reverse transcription-PCR (RT-PCR) analyses. For northern blot analysis, total RNA was extracted from 5^{th} instar larval midgut tissue and midgut-free larval carcass using TRIzol reagent (Invitrogen) and separated on 1% agarose gels in 10 mM sodium phosphate buffer (Pelle and Murphy, 1993). RNA was then transferred onto Hybond-XL nylon membranes (Amersham Pharmacia Biotech., Piscataway, NJ, USA) with 10xSSC [3 M NaCl, 0.3 M trisodium citrate·$2H_2O$ in diethyl pyrocarbonate (DEPC)-treated dH_2O, pH 7.0] using a positive pressure transfer apparatus (Stratagene, La Jolla, CA, USA) and cross-linked with UV light using a Stratalinker (Stratagene). cDNA was labelled with ^{32}P-dCTP using the Random Prime Labeling kit (Invitrogen) to produce probes for hybridization.

Blots were prehybridized, hybridized, and washed at 65ºC according to Church and Gilbert (1984).

RT-PCR analyses were conducted to examine the level of gene expression in different developmental stages or in different tissues. Midgut tissues were pooled from 3 to 10 4th instar feeding larvae or larvae molting from 4th to 5th instar. The feeding stage was divided into early (12-24 h after molting), mid (36-48 h after molting) and late (60-72 h after molting) whereas the molting stage was divided into Phase I (premolt, the first 6-8 h period after larvae stop feeding and their cuticle turns green), Phase II (midmolt, the 3-5 min period when shedding of the old cuticle reached the midpoint of the body) and Phase III (postmolt, the 1h period after ecdysis before larvae recommence feeding) at 21±1ºC. To examine the tissue distribution of gene expression foregut, hindgut, Malpighian tubules, fat body, tracheae and integument were dissected from 20-30 feeding 5th instar larvae and subjected to RT-PCR analysis. Total RNA was extracted from corresponding tissues using TRIzol reagent and resuspended in 100 µl of DEPC water. The RNA was treated with DNase1 (Invitrogen) and RNase Stop (Invitrogen) for 15 min at room temperature and then heated to 65^{o}C for 10 min. RNA samples were extracted with phenol and then with phenol/chloroform (1:1), the RNA precipitated with ethanol and resuspended in 100 µl DEPC-treated water. RT-PCR was performed with 100 ng of total RNA using the SuperScript One-Step RT-PCR kit (Invitrogen) in a 25 µl reaction volume and the amount of RNA was adjusted so that the amplification of the *M. configurata tubulin* (*McTUB)* and *actin* (*McACT)* was the same in all samples. The RT-PCR conditions were: 50ºC for 30 min, 94ºC for 2 min, followed by 18-40 cycles of 94ºC for 15 sec, 55ºC for 30 sec, 70ºC for 1 min, and a final extension at 72ºC for 10 min. In general, 28 cycles were used unless stated otherwise. RT-PCR products were separated on 1% agarose gels.

2.5 Expression of Proteins in *E.coli*

Protein coding regions without the signal peptide were amplified using the One-Step RT-PCR kit (Invitrogen). The amplified products were purified after separation in an agarose gel and cloned into the pGEM-T Easy vector. The fragments were excised by digestion with *Nde*I and *Xho*I and cloned into the pET28 bacterial expression vector (Novagen Inc., Madison, WI, USA) in three different *E. coli* strains: BL21 (DE3), Rosetta 2 (DE3) and Rosetta gami 2 (DE3) pLysS to test for optimal expression.

E. coli cells were grown in 100 ml of Luria broth to an OD_{600} of 0.6, at which time protein expression was induced with the addition of IPTG (1 mM) and the cultures incubated at 37°C for an additional 5 h. The cultures were centrifuged at 4000 rpm for 10 min and the cell pellets resuspended in 1 ml of 0.5 M NaCl, 20 mM Tris-HCl (pH 8.0). The cells were lysed using a French press, 5 µl of 10 mg/ml deoxyribonuclease I and 2.5 µl of 100 mg/ml ribonuclease A added and the mixture incubated on ice for 15 min. Finally, 40 µl of 25X protease inhibitor cocktail (Roche Diagnostics GmbH, Indianapolis, IN, USA) was added and the suspension centrifuged at 4000 rpm for 10 min to collect the soluble fraction. The pellet was resuspended in 0.5 M NaCl, 20 mM Tris-HCl (pH 8.0), 6 M urea, and incubated on ice for 1 h. The suspension was centrifuged and the supernatant collected as the inclusion body fraction.

2.6 Extraction of Proteins from Insect Tissues

Peritrophic matrices (PMs) and midgut tissues were dissected separately from 20 mid 5^{th} instar larvae and immediately washed in a droplet of ice cold protease inhibitor cocktail (Roche Diagnostics GmbH) to remove food contents. The PM or midgut tissues were then transferred into ice cold Ringer's solution (153 mM NaCl, 2.68 mM KCl, 1.36 mM $CaCl_2 \cdot 2H_2O$) containing protease inhibitor cocktail. Fifteen PM or midguts were ground in a 1.5 ml microcentrifuge tube using a pestle and washed three times with 200 µl of Ringer's solution with protease inhibitor cocktail. After each wash, the material was recovered by centrifugation at 20,800 g at 4°C for 1 min and the supernatant representing the "loosely associated" proteins collected and pooled. This fraction may contain some remnants of the food bolus as well as genuine PM associated proteins. The "strongly associated" proteins were extracted from the remaining PM pellet by grinding vigorously in a protein extraction solution of 2.5% SDS, 5% β-mercaptoethanol with or without 500 mM NaCl followed by incubation at 100°C for 5 min and the supernatant collected after centrifugation at 20,800 g for 5 min.

2.7 One Dimensional and Two Dimensional Protein Separations

Denaturing one dimensional (1D) gel electrophoresis (Laemmli, 1970) was performed by boiling samples in SDS loading buffer containing 50 mM Tris (pH 6.8), 2% SDS, 0.05% bromophenol blue, 10% glycerol, and 2% β-mercaptoethanol. For separation of proteins under nondenaturing conditions the protocol described in Hegedus et al. (2003) was used. Briefly,

samples were mixed with loading buffer without β-mercaptoethanol on ice and the electrophoresis was conducted at 4°C. Proteins were separated in 12% polyacrylamide gels containing either 1% SDS (denatured proteins) or 0.5% SDS (nondenatured proteins) at 70V for 200 min and stained with Coomassie blue (Wong et al., 2000).

Two dimensional (2D) gel electrophoresis was performed using 7 cm immobilized pH gradient strips (BioRad Laboratories, Hercules, CA, USA). The protein samples from the loosely or strongly associated PM preparations were dialyzed using a 3500 MW cut off dialysis cassette (Pierce, Rockford, IL, USA) against five changes of 10 mM Tris-HCl (pH 8.0) at 4°C over 48 h. The strips were rehydrated overnight at 4°C in a solution of 6.7 M urea, 2% 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS), 2 mM tributyl phosphine (TBP), 0.5% ampholytes (pH 3-10) and 0.001% bromophenol blue dye containing 3-9 µg protein. The first dimension was focused in three stages: 1) held at 250 V for 30 min, 2) held at 4000 V for a total of 9000 V-h, and 3) held at 500 V until used. Strips were equilibrated by washing in a solution of 6 M urea, 2% SDS, 375 mM Tris-HCl (pH 8.8), 20% glycerol, 135 mM TBP for 10 min and then in a solution of 6 M urea, 2% SDS, 375 mM Tris-HCl (pH 8.8), 20% glycerol, and 135 mM iodoacetamide. Proteins were separated in the second dimension in 12% polyacrylamide gels at 70 V and stained using the "Silver Quest Kit" (Invitrogen).

2.8 Antisera Production and Western Blot Analysis

Antisera were generated using conjugated peptides or recombinant proteins. Peptide-derived antisera were developed by conjugating two synthetic peptides, each of 20 amino acids, synthesized at Sheldon Biotechnology Centre, McGill University (Montréal, QC, Canada) or EzBioLab Company (Carmel, IN, USA). Peptides were conjugated to bovine serum albumin using the Imject Activated Immunogen Conjugation Kit (Pierce) or to keyhole limpet (*Megathura crenulata*) hemocyanin (EzBioLab). In all antisera development, New Zealand White rabbits were injected with 100-120 µg of total protein in Titremax adjuvant (Sigma), followed by two subsequent injections at 2-week intervals. The corresponding antisera were tested against available recombinant proteins. Preimmune sera were used as controls for all immunological experiments.

After 1D gel electrophoresis, proteins were transferred to a nitrocellulose or PVDF membrane (Amersham Pharmacia Biotech) in Towbin transfer buffer (Towbin et al., 1979) with

(denatured proteins) or without (nondenatured proteins) 20% methanol. The membranes were blocked using 5% skim milk powder in PBS-T buffer (137 mM NaCl, 2.7 mM KCl, 10.1 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 0.2% Tween 20, pH 7.5) overnight at 4°C. After blocking, the membranes were incubated with primary antisera and then with horseradish peroxidase labelled antirabbit IgG for 1 h each at room temperature. The location of the secondary antibody was determined using the chemiluminescence detection method with Immobilon western detection reagents (Millipore, Billerica, MA, USA). Antihexahistidine (Santa Cruz Biotechnology, Santa Cruz, Ca, USA), antiphosphothreonine and antiphosphotyrosine (Calbiochem, La Jolla, CA, USA) antibodies were used according to the manufacturer's recommendations.

2.9 Mass Spectrometry of PM Associated Proteins

Protein mass spectrometry was conducted under contract to the National Research Council of Canada (Saskatoon, SK, Canada). Protein bands or slices were excised from Coomassie blue stained 1D gels whereas protein spots were excised from silver stained 2D gels using a Proteome Works 2D spot cutter (Bio-Rad Laboratories) and placed in a 96-well microtitre plate (Sigma). The proteins were then automatically destained, reduced with dithiothreitol, alkylated with iodoacetamide, and digested with porcine trypsin (sequencing grade, Promega) using a MassPREP protein digest station (Micromass, Pointe-Claire, QC, Canada). The digest was evaporated to dryness, then dissolved in 12 µl of 1% aqueous trifluoroacetic acid, of which 6 µl was analyzed by liquid chromatography tandem-mass spectrometry (LC-MS/MS) using a capLC ternary HPLC system (Waters, Milford, MA, USA) interfaced to a Q-ToF Ultima Global hybrid tandem mass spectrometer fitted with a Z-spray nanoelectrospray ion source (Micromass). Solvent A comprised 0.2% formic acid in water, while solvent B consisted of 0.2% formic acid in acetonitrile. The peptide digest sample was loaded onto a C18 trapping column (Symmetry 300, 0.35 x 5 mm Opti-pak; Waters) and washed for 3 min using solvent A at a flow rate of 30 µl/min. The flow path was then switched using a 10-port rotary valve and the sample eluted onto a C18 analytical column (PepMap, 75 µm x 15 cm, 3-µm particle size; LC Packings). Separations were performed using a linear gradient of 5:95% to 60:40 % B:A over 43 min. The composition was then changed to 80:20 % B:A and held for 10 min to flush the column before re-equilibrating for 7 min at 0:100 % A:B. Mass calibration of the Q-ToF instrument was performed using a product ion spectrum of Glu-fibrinopeptide B acquired

over the *m/z* range 50 to 1900. LC-MS/MS analysis was carried out using data dependent acquisition, during which peptide precursor ions were detected by scanning from *m/z* 400 to 1900 in ToF MS mode. Multiply charged (2+, 3+, or 4+) ions rising above predetermined threshold intensity were automatically selected for ToF MS/MS analysis, a process that involves charge state recognition and transfer of the selected (precursor) ions to a collision cell, where they undergo collision induced dissociation in the presence of argon to form fragment (product) ions. Product ion spectra were acquired over the *m/z* range 50 to 900 by varying the collision energy and charge state recognition. LC-MS/MS data were processed using ProteinLynx v2.15 software (Micromass) set to default parameters and searched against an in-house *M. configurata* midgut EST database using MASCOT v2.2 (Matrix Science Inc., Boston, MA, USA). Searches were performed using a fragment ion mass tolerance of 0.3-0.6 Da, one missed cleavage, carbamidomethylation of cysteine as the fixed modification and oxidation of methionine as the variable modification. Database search results were considered positive when the probability based mowse score exceeded the 95% confidence threshold.

2.10 Prediction of Protein Features

Putative signal peptides were identified using SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP/) and *N*-linked or *O*-linked glycosylation sites were identified using NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/) or NetOGlyc3.1 (Julenius et al., 2005) (http://www.cbs.dtu.dk/services/NetOGlyc/), respectively. Serine, threonine and tyrosine phosphorylation sites were predicted using NetPhos 2.0 (www.cbs.dtu.dk/services/NetPhos/) (Blom et al., 1999). The proteins were also scanned for glycosylphosphatidylinositol (GPI) membrane anchor sequence using big-PI Predictor (mendel.imp.ac.at/sat/gpi/gpi_server.html) (Eisenhaber et al., 2000) and transmembrane domains using TMHMM Server v. 2.0 based on the hidden Markov model (www.cbs.dtu.dk/services/TMHMM-2.0/). The hydrophobicity-hydrophilicity of the proteins was determined using Hydrophobicity plot 1.0 (www.bmm.icnet.uk/~offman01/hydro.html) or ProtScale (Gasteiger et al., 2005) (http://www.expasy.ch/cgi-bin/protscale.pl).

3. MORPHOLOGY OF THE PERITROPHIC MATRIX, SYNTHESIS AND DEGRADATION OF CHITIN ASSOCIATED WITH THE PERITROPHIC MATRIX[1]

3.1 Introduction

Insect PMs have been classified into two types according to their site of production (Wigglesworth, 1930; Waterhouse, 1953a; b). The Type I PM is produced along the entire length of the midgut epithelium, which is the case in most lepidopteran larvae and dipteran adults. The Type II PM is produced by a special ring of cells or an organ called the cardia in the foregut-midgut boundary; the former is found in some lepidopteran larvae and the latter is typical of dipteran larvae. The difference in the site of PM synthesis also causes some differences in PM morphology. Type I PMs generally delaminate from the midgut epithelium to form multilayered structures (Lagermalm et al., 1950). In contrast, Type II PMs are extruded from the anterior midgut and lack obvious layers; however, they may contain several fused laminae (Waterhouse, 1954). For both types of PM, chitin and proteins containing peritrophin A domains (PAD) that serve a chitin binding function are the key structural components. The localization and timing of the synthesis of these components are thus critical to our understaning of the formation and morphology of the PM.

Chitin, a polymer of β-(1, 4)-*N*-acetyl-D-glucosamine (GlcNAc), is the major structural polysaccharide of the cuticular lining in the foregut, hindgut, tracheae, integument and the peritrophic matrix (PM) in insects (Gillott, 2005). The chitin associated with the PM is synthesized by the midgut epithelial cells as microfibrils with diameters of 2-10 nm and lengths exceeding 500 nm (Lehane, 1997). Ten or more microfibrils are bundled together, mostly in a parallel fashion, to form the primary mesh-like network of the PM. PMs may appear as lattice-like structures organized into orthogonal or hexagonal networks (Ryerse et al., 1992; Harper and Hopkins, 1997), or as random felt-like structures of fibers (Adang and Spence, 1981).

Chitin synthesis and degradation are controlled by several enzymes, primarily chitin synthase (CHS), chitinase (CHI) and β-*N*-acetylglucosaminidase (NAG) (Merzendorfer and Zimoch, 2003). Two chitin synthases are found in most insects; CHS-B, which is produced in the midgut and involved in PM biosynthesis, and CHS-A, which is responsible for the synthesis of chitin in the integument (Coutinho and Henrissat, 1999; Merzendorfer, 2006). Both enzymes catalyze the transfer of GlcNAc to the growing chitin chain using uridine diphosphate-*N*-

[1] Part of this chapter is published in *Insect Molecular Biology* **19:**163-175, 2010.

acetylglucosamine (UDP-GlcNAc) as the donor (Martinez et al., 2009). Chitin is digested by enzymes that cleave either internally (endochitinases) or at the end of the chitin chain (exochitinases). CHI is an endochitinase and hydrolyses internal bonds in polymeric (insoluble) chitin, forming smaller, soluble chito-oligosaccharides (Arakane and Muthukrishnan, 2010). The chito-oligosaccharides are further hydrolysed to 2-acetamido-2-deoxyglucopyranoside by the exochitinase NAG (Fukamizo and Kramer, 1985). Numerous chitinases have been described from insects and differentiation of these proteins is less clear relative to the distinction between CHS enzymes. CHI genes may be expressed in multiple tissues, including the integument and midgut, and all share a high degree of sequence similarity (Arakane and Muthukrishnan, 2010). Biochemical studies on both CHS and CHI revealed that both enzyme groups are zymogenic, requiring proteolytic processing by serine proteases for activation (Mayer et al., 1980; Koga et al., 1989; Shahabuddin et al., 1996).

During insect development, the PM must be degraded periodically and replaced with a new PM to allow for growth and maturation, as in the case of periodic ecdysis of old cuticular integument (Kramer and Koga, 1986). This requires the coordinated action of chitin synthase and chitinolytic enzymes, which was proposed to be accomplished at the level of gene transcription (Bolognesi et al., 2005). This model involves coordinated expression of the genes encoding CHS-B and CHI during development. Thus, *CHS-B* is downregulated during the molt and upregulated during feeding, whereas *CHI* is expressed during the molt (Arakane et al., 2004; Bolognesi et al., 2005; Zimoch et al., 2005). However, these studies did not examine *CHS-B* and *CHI* expression in physiologically different developmental stages. For example, molting larvae were examined only at a single stage, although molting comprises multiple steps including; i) a dormant period in which the larvae cease feeding and remain sessile (referred to molting Phase I in the current study), ii) shedding of old cuticle (molting Phase II), and iii) the nonfeeding period prior to the resumption of feeding (molting Phase III).

The current study first examines the physical architecture of the PM in feeding, molting and starved larvae, followed by characterization of genes encoding the *M. configurata* CHS-B (McCHS-B), CHI (McCHI) and NAG (McNAG) enzymes. Specifically, the expression of *McCHS-B*, *McCHI* and *McNAG* was examined in various tissues, as well as in feeding (early, middle, late), molting (Phase I, Phase II, Phase III) and starved larvae. The ortholog of *M. sexta* chymotrypsin like protease (MsCTLP1), which is the putative *CHS-B* activator in *M. sexta*

(Broehan et al., 2007), was also identified in *M. configurata* and its expression pattern was examined.

3.2 Material and Methods

3.2.1 Microscopy

PMs were examined by epifluorescence, confocal, and scanning electron microscopy (SEM). Epifluorescence microscopy was used to image the general physical architecture of the PM at different conditions. PMs from feeding or starved (24 h without food) 5^{th} instar larvae and molting Phase II larvae (5^{th} to 6^{th} instar) (molting phases were defined in Chapter 2) were incubated in 0.5% (w/v) Fluorescent Brightener 28 (FB28) (Sigma, St Louis, MO, USA). The PMs were rinsed in distilled water and photographed under brightfield or fluorescence microscopy (Zeiss Axiocam) with a 0.8X objective (Neolumar S) using a digital camera (Axiocam HRc).

Confocal microscopy was used to examine the PM in greater detail to investigate the multilayered structure observed using FB28 staining with fluorescent miscroscopy. PMs from 4^{th} and 5^{th} instar feeding larvae were incubated in 10 µg/ml Alexafluor 488 conjugated wheat germ agglutinin (AWGA) (Invitrogen, Carlsbad, CA, USA) in phosphate buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10.1 mM Na_2HPO_4, 1.8 mM KH_2PO_4) for 30 min and then stored in PBS at 4°C in the dark until examined by confocal microscopy. Confocal microscopy was conducted using a Zeiss LSM 710 confocal microscope fitted with a 20X0.8 numerical aperature objective lens. Zeiss ZENN 9.0 software was used to control the microscope. The Alexafluor 488 was excited using either an Argon LGN3001 488 nm laser with the pinhole set at 1 Airy unit or a Coherent Chameleon laser tuned to 720 nm with the pinhole wide open. In either case, emissions from 494-645 nm were collected, with a pixel dwell of 6.30 µs and line averaging set to 2. Voxel size was between 0.4x0.4x3.0 µm and 0.5x0.5x3.0 µm (XYZ). Photomultiplier tube gain and laser power were increased to maintain image brightness as imaging depth increased. Confocal images were resolved using Autoquant X version 2.2.0 (Autodeblur Gold CF package) using the default settings for three dimensional decovolution. Adjacent Z-stacks were then stitched using XUV tools (http://www.xuvtools.org) (Emmenlauer et al., 2009) and visualized using Imaris 7.1.1 software (http://www.bitplane.com/go/products/imaris). Full resolution TIFF files were captured showing the XZ plane to get a cross-section of the intact PM.

For SEM analyses, PMs were dissected from middle stage 5th instar *M. configurata* or *Manduxa sexta* larvae, the latter provided by Dr. Andrew Keddie (University of Alberta), for comparison of the PMs. PMs were fixed for 30 min in 3% glutaraldehyde in 0.1 M cacodylate buffer (pH 7.3) at room temperature and then rinsed in cacodylate buffer for 30 min. Following washing, the samples were postfixed in 2% OSO_4 in cacodylate buffer at 4°C for 2 h. PMs were rinsed again in cacodylate buffer for 30 min and then dehydrated in increasing acetone concentrations (15%, 30%, 50%, 70%, 90%, and 100%) of 30 min each. Specimens were dried using a critical point apparatus according to the method of Cohen et al. (1968) and then coated with gold particles.

3.2.2 Identification of the Genes Encoding Midgut Chitin Synthase and Chitinolytic Enzymes

3.2.2.1 Chitin synthase-B (McCHS-B)

Lepidopteran *CHS-B* cDNA sequences from *M. sexta* (*MsCHS-B*, GenBank accession number, AAX20091), *Spodoptera exigua* (*SeCHS-B*, ABI96087), *S. frugiperda* (*SfCHS-B*, AAS12599), and *Ostrinia furnacalis* (*OfCHS-B*, ABX46067) were used to search an *M. configurata* EST database using BLAST. A partial cDNA (approximately 2.2 kb) that lacked a region encoding the amino terminus of the enzyme was obtained. 5' RACE failed to generate the missing 5' information; therefore, a RT-PCR based approach was conducted using total midgut RNA from the 5th instar *M. configurata* larvae fed on *Brassica napus* leaves as described in Chapter 2. In the RT-PCR, a forward primer (McCHS-B Fp1, Table 3.1) that corresponded to a conserved region in *MsCHS-B*, *SeCHS-B*, *SfCHS-B*, and *OfCHS-B* that was lacking in the 5' sequence in *McCHS-B*, and a gene specific reverse primer (McCHS-B Rp1, Table 3.1) within the partial *McCHS-B* cDNA were used. The RT-PCR products were separated on a 1% agarose gel and a 1.1 kb band excised from the gel, purified, and cloned into the pGEM-T Easy (Promega, Madison, WI, USA) vector. This product captured additional McCHS-B sequence; however, approximately 1.2 kb of 5'cDNA sequence remained. To obtain this information, the 1.1 kb fragment was labelled with ^{32}P-dCTP and used to screen a *M. configurata* bacterial artificial chromosome (BAC) genomic DNA library (Hegedus et al., 2008). BAC DNAs from positive clones were extracted using the Large-Construct Kit (Qiagen, Valencia, CA, USA) and verified with PCR to contain the *CHS-B* gene. BAC DNAs were sequenced using a 454 GS-FLX

pyrosequencer (Roche Applied Science, Branford, CT, USA) at the National Research Council of Canada, Saskatoon. To identify the exons, the BAC sequence was aligned with the cDNAs from *MsCHS-B*, *SeCHS-B*, *SfCHS-B* and *OfCHS-B*. A forward primer (McCHS-B Fp2, Table 3.1) complementary to a region starting with the putative start codon was designed. A final RT-PCR reaction was conducted using McCHS-B Fp2 and a reverse primer (McCHS-B Rp2, Table 3.1) designed within the 1.1 kb fragment of *McCHS-B* to obtain the missing 5' cDNA sequence and to verify predicted intron-exon splice junctions.

3.2.2.2 Chitinase (McCHI)

A cDNA encoding the *M. configurata* midgut CHI ortholog was not found in the midgut EST database. Therefore, sequences of midgut chitinase cDNAs from *M. brassicae* (*MbCHI*, ACL30984), *S. exigua* (*SeCHI*, ADI24346), *S. frugiperda* (*SfCHI*, AAS18266) and *S. litura* (*SlCHI*, BAB12678) were used to design conserved forward and reverse primers. An RT-PCR reaction was conducted using total combined midgut RNA from Phase I, Phase II, and Phase III larvae molting from 5^{th} to 6^{th} instar (as described in Chapter 2) with the conserved primers (McCHI Fp1 and McCHI Rp1, Table 3.1). The RT-PCR product was separated on a 1% agarose gel, cloned into pGEM-T Easy (Promega) and sequenced.

The missing sequences from both 5' and 3' were obtained using the GeneRacer™ Advanced RACE Kit (Invitrogen). In 5' RACE analysis, target mRNA was converted to cDNA using McCHI Gsp1(Table 3.1), which was amplified with a forward primer provided by the kit and McCHI Gsp2 (Table 3.1). In 3' RACE analysis, a gene specific forward primer (McCHI Gsp3) (Table 3.1) and a universal poly-A tail specific reverse primer were used. Other details regarding RACE analysis were explained in Chapter 2. The RACE products were separated by electrophoresis in 1% agarose gels, cloned into pGEM-T Easy (Promega) and sequenced.

3.2.2.3 β-N-Acetylglucosaminidase (McNAG)

A cDNA encoding a NAG ortholog was not found in the *M. configurata* midgut EST database; therefore, sequences of midgut NAGs from *Agrotis ipsilon* (AiNAG, GU985280), *Bombyx mori* (BmNAG, NM_001044002), *Choristoneura fumiferana* (CfNAG, DQ005717), *M. sexta* (MsNAG, AY368703), *O. furnacalis* (OfNAG, DQ887769), and *Xestia c-nigrum* (XnNAG, FJ848784) were used to design conserved primers. An RT-PCR (as described in

Chapter 2) was conducted using total midgut RNA extracted from larvae in molting Phase III and the McNAG Fp1 and McNAG Rp1 primers (Table 3.1). The RT-PCR product was separated by electrophoresis in a 1% agarose gel, cloned into pGEM-T Easy (Promega) and sequenced.

Table 3.1 Primers used to obtain cDNA sequences of *M. configurata McCHI*, *McCHS-B* and *McNAG*.

Primer	**Sequence**
McCHI Gsp1	CATTCAGGCGTCGGAGTAGTGTTTC
McCHI Gsp2	TGAGCGAGGTGGTGGTACAGTGTAG
McCHI Gsp3	TCTGGTTGCTCGGTGCCTTGATTCTG
McCHI Fp1	CTGGGCTATTGACATGGATGACTTC
McCHI Rp1	ATACGACAATATTGTTCAGTAAC
McCHS-B Fp1	CAAGAAACAAAAGGATGGAA
McCHS-B Rp1	CCGAACATCAACAGACCCTGATACCA
McCHS-B Fp2	ATGGCGACGAAGCCCAAGAC
McCHS-B Rp2	AGTAGTCAGGATCAACAATTC
McNAG Fp1	GACATGTTCCACATGGGCGGTGACGAGGT
McNAG Rp1	CCTTCGTTCTGGTAGCACCACTC

Abbreviations: CHI, chitinase; CHS-B, chitin synthase-B; Fp, forward primer; NAG, β-*N*-acetylglucosaminidase; Rp, reverse primer; GSP, gene specific primer; Mc, *M. configurata*.

3.2.2.4 Ortholog of putative chitin synthase activator, MsCTLP1

A full length cDNA encoding an ortholog of *M. sexta* chymotrypsin like protease (MsCTLP1), *M. configurata* McSP33, was identified in the *M. configurata* midgut EST database.

3.2.3 Analysis of Gene Expression

Expression of target genes was examined using RT-PCR or quantitative real-time PCR (qPCR). RT-PCR analysis was used to examine the tissue specific expression of *McCHS-B*, *McCHI*, *McNAG* and *McSP33* using the gene specific primers (Table 3.2) in foregut, midgut, hindgut, Malpighian tubules, integument, tracheae and fat body from 5th instar larvae as described in Chapter 2. In addition, the expression levels of *McNAG* and *McSP33* in early,

middle and late stage feeding or 24 h starved 5th instar larvae, or Phase I, Phase II, and Phase III larvae molting from 5th to 6th instar were also examined by RT-PCR.

qPCR analysis was used to compare the expression of *McCHI* and *McCHS-B* throughout development. In qPCR analysis, total RNA from early, middle, late stage feeding and 24 h starved 5th instar larvae and Phase I, Phase II, and Phase III larvae molting from 5th to 6th instar was purified using the Illustra RNAspin mini kit (GE Healthcare, Little Chalfont, UK). Single strand cDNA was synthesized from approximately five micrograms of purified total RNA using the Superscript First-Strand Synthesis System (Invitrogen). Amplification of the *M. configurata* actin (*McACT*) gene was used as a control to ensure equivalent cDNA template in samples. Additionally, an initial qPCR was conducted using the SYBR Green qPCR kit (Finnzymes Oy, Espoo, Finland) with middle stage feeding and molting Phase II cDNA with 0.25 µM of each primer to test the efficiency of the primer binding. Only primers with efficiencies from 90 to 110% were used for qPCR analysis. cDNAs were diluted 1:100 and 6 µl used as a template in a 20 µl qPCR reaction under the following conditions: 95°C for 10 min, followed by 40 cycles of 95°C for 10 sec, 60°C for 30 sec and 72°C for 30 sec. A melting curve of denatured double stranded amplicon was established to test the purity of the products using the following conditions: 95°C for 15 sec, 60°C for 1 min, 0.3°C increments for 15 sec each until reaching 95°C. Data were obtained from each of two biological replicates with three technical replicates.

3.2.4 *β-N-Acetylglucosaminidase Activity*

Exochitinase NAG activity was examined in the PMs from early, middle, late stage feeding and starved (24 h) 5th instar larvae as well as Phase I, Phase II, and Phase III larvae molting from 5th to 6th instar using the chitinase assay kit (Sigma). PMs were dissected from five larvae, combined and ground with a pestle in 100 µl ddH_2O. Samples were centrifuged at 14,000 rpm in a microcentrifuge and the supernatant used for the enzymatic reactions according to the manufacturer's recommendations. Briefly, 1-10 µl of the PM extract was incubated with 90-99 µl of the exochitinase substrate, 4-Nitrophenyl *N*-acetyl-β-D-glucosaminide at 37ºC for 30 min. The reaction was stopped by adding 200 µl of Na_2CO_3 stop solution. The A_{405} nm was determined as a measure of substrate hydrolysis, which releases 4-nitrophenol. The data were subjected to analysis using one-way ANOVA (P=0.05) in the Statistica software (Statistica, 1997). When a

significant difference was detected, a Tukey HSD test (P=0.05) was used to determine which means were significantly different.

Table 3.2 Primers used in gene expression analysis of *M. configurata McCHI*, *McCHS-B* and *McNAG* and *McSP33* by RT-PCR or qPCR.

Primer[1]	**Sequence**
McACT Fp	GATCGTGCGTGACATCAAGG
McACT Rp	CACACTTCATGATCGAGTTG
McCHI Fp	GAGGGAGAGCCGATAGTCACG
McCHI Rp	TCAGTTGGCGGTTCAGGAG
McCHS-B Fp	CCTCATCTTCGGCTCGTTTTTCA
McCHS-B Rp	TGTTTTCTTCTGCGCCACTTCC
McNAG Fp	GACATGTTCCACATGGGCGGTGACGAGGT
McNAG Rp	CCTTCGTTCTGGTAGCACCACTC
McSP33 Fp	CAATGTCACCCACCCCAACTA
McSP33 Rp	AGACAAAGCACCGCCAGACC
McTUB Fp	AGCGTACCATCCAGTTCGTG
McTUB Rp	TGGCGTACATGAGGTCGAAC

[1]**Abbreviations:** ACT, actin; CHI, chitinase; CHS-B, chitin synthase-B; Fp, forward primer; Mc, *M. configurata*; NAG, β-*N*-acetylglucosaminidase; Rp, reverse primer; SP, serine protease; TUB, tubulin.

3.3 Results

3.3.1 Physical Architecture of M. configurata *PM during Development*

The PMs from larvae at three developmental stages were found to have different morphological features. Bright field and epifluorescence microscopy showed that the PM from feeding and starved larvae was translucent; however, dissection of the PM from starved larvae showed it to be more fragile and about 1.5 to 2X longer than the PM from feeding larvae (Figure 3.1). Treatment with FB28 revealed additional structural details. The PM from feeding larvae had numerous, closely spaced striations, presumably bundles of chitin strands, running perpendicular to the long axis of the PM. Longitudinal or angular bundles were rarely observed, except for two major bundles running in parallel along the length of the PM. The striations in the PM from starved larvae were disorganized and less evident than those in feeding larvae. The PM

consisted of an inner opaque layer and an outer translucent layer in molting larvae. The inner opaque material had vertical striations as seen with the PM from feeding and starved larvae, while the outer material was amorphous and stained less intensely with FB28, suggesting the latter PM is an immature, newly synthesized PM. It is noteworthy that a PM was either not found or much smaller in molting Phase I larvae (data not shown).

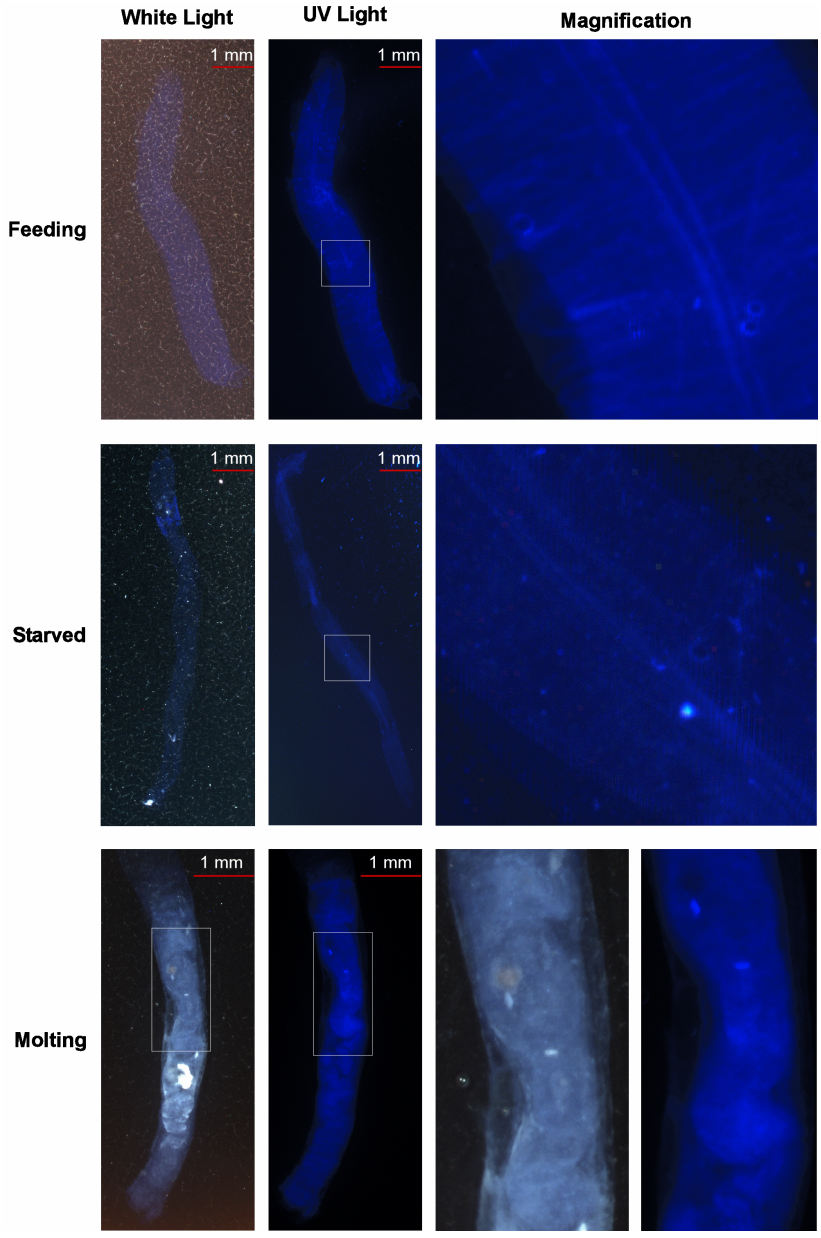


Figure 3.1 Peritrophic matrices (PMs) from *Mamestra configurata* feeding, starved and molting (from 5th to 6th instar) larvae. PMs were viewed under white or UV light after treatment with the chitin binding fluorescent brightener FB28 using epifluorescence microscopy. Magnification of the PM from regions noted (boxes) is shown at the right.

The PMs from all three stages studied consisted of at least two layers (Figure 3.1). Treatment with AWGA followed by confocal microscopy revealed that the PM has two major layers, an internal layer covering the bolus and an external layer covering the whole PM (Figure 3.2). Between these two major layers, 2-4 and 8-10 thinner layers were detected in 4th and 5th instar larvae, respectively (Figure 3.2).

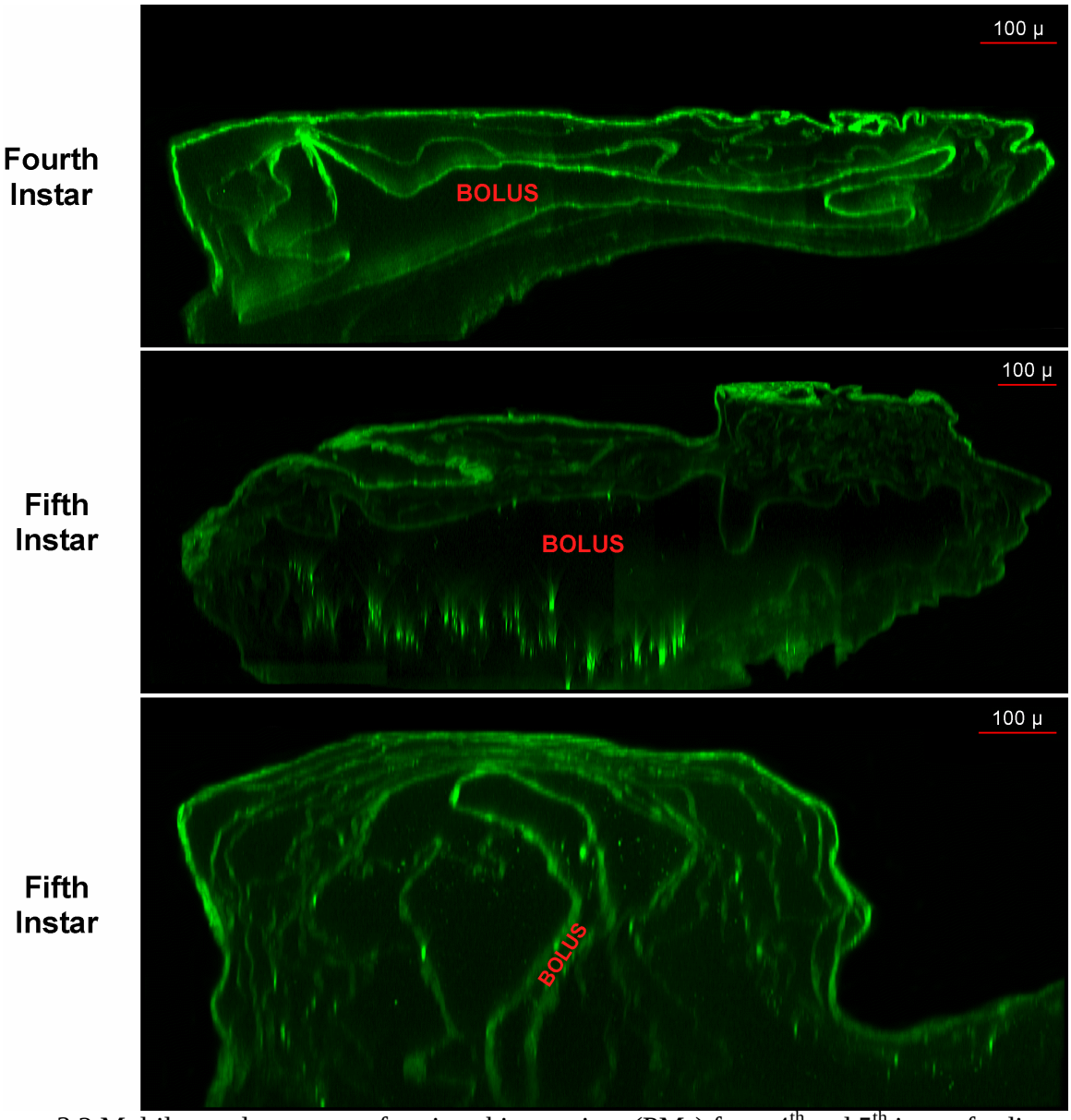


Figure 3.2 Multilayered structure of peritrophic matrices (PMs) from 4th and 5th instar feeding *Mamestra configurata* larvae. PMs were viewed using a confocal microscope after treatment with Alexafluor 488 conjugated to Wheat Germ Agglutinin (AWGA). The location of the bolus (endoperitrophic space) is indicated.

SEM analysis of PMs from *M. configurata* and *M. sexta* revealed a continuous texture with numereous wrinkles, on both the luminal and epithelial sides (Figure 3.3). Neither a mesh-like structure nor pores were visible at this magnification.

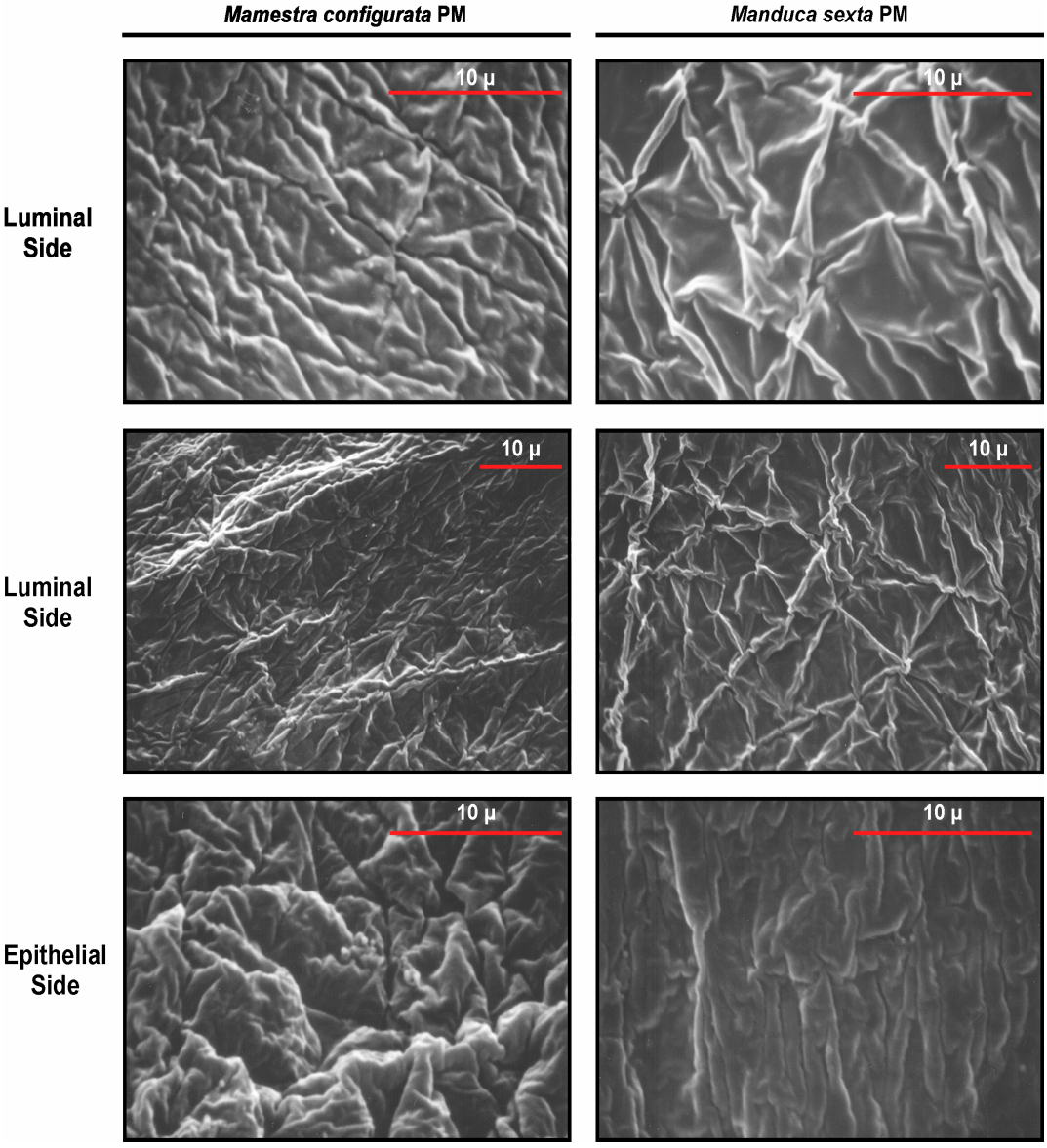


Figure 3.3 Scanning electron microscopy of peritrophic matrices (PMs) from middle stage feeding 5th instar *Mamestra configurata* and *Manduca sexta* larvae.

3.3.2 Characterization of the Genes Encoding Midgut Chitin Synthase and Chitinolytic Enzymes

3.3.2.1 Chitin synthase (McCHS-B)

A search of the *M. configurata* midgut EST database for a *CHS-B* ortholog revealed one partial cDNA of 2365 bp including the poly-A tail. The EST encoded the carboxy terminal 675

amino acids (Figure 3.4) including part of the catalytic domain (Marchler-Bauer et al., 2009). The sequence of the entire open reading frame was derived using a combination of RT-PCR and analysis of *M. configurata* BAC clone. RT-PCR using an upstream primer complementary to a region conserved among lepidopteran CHS-B genes revealed an additional 936 bp fragment encoding 312 amino acids (Figure 3.4). This fragment was used to identify a BAC clone that provided an additional 1665 bp of 5' sequence information corresponding to 555 amino acids. McCHS-B was predicted to be 1542 amino acids long with a predicted molecular weight of 175.4 kDa (Figure 3.4).

McCHS-B lacked a signal peptide and was predicted to have 16 transmembrane helices, 9 of which were upstream of the central catalytic domain located at amino acids 631-886. The McCHS-B catalytic domain exhibited a high degree of similarity with that of other lepidopteran CHS-B enzymes, ranging from 94.9 to 96.1% identity when compared to SeCHS-B, SfCHS-B, MsCHS-B, and OfCHS-B (Figure 3.4). Comparison of the entire CHS-B sequences revealed that McCHS-B was most similar to SeCHS-B (77.2% identity), followed by SfCHS-B (75.7%), MsCHS-B (72.3%), and OfCHS-B (71.7%). All five CHS-Bs had typical CHS signature motifs (Hogenkamp et al., 2005), including “EDR” (positions 838-840), “QRRRW” (875-879), and “WGTRE” (1058-1062).

3.3.2.2 Chitinase (McCHI)

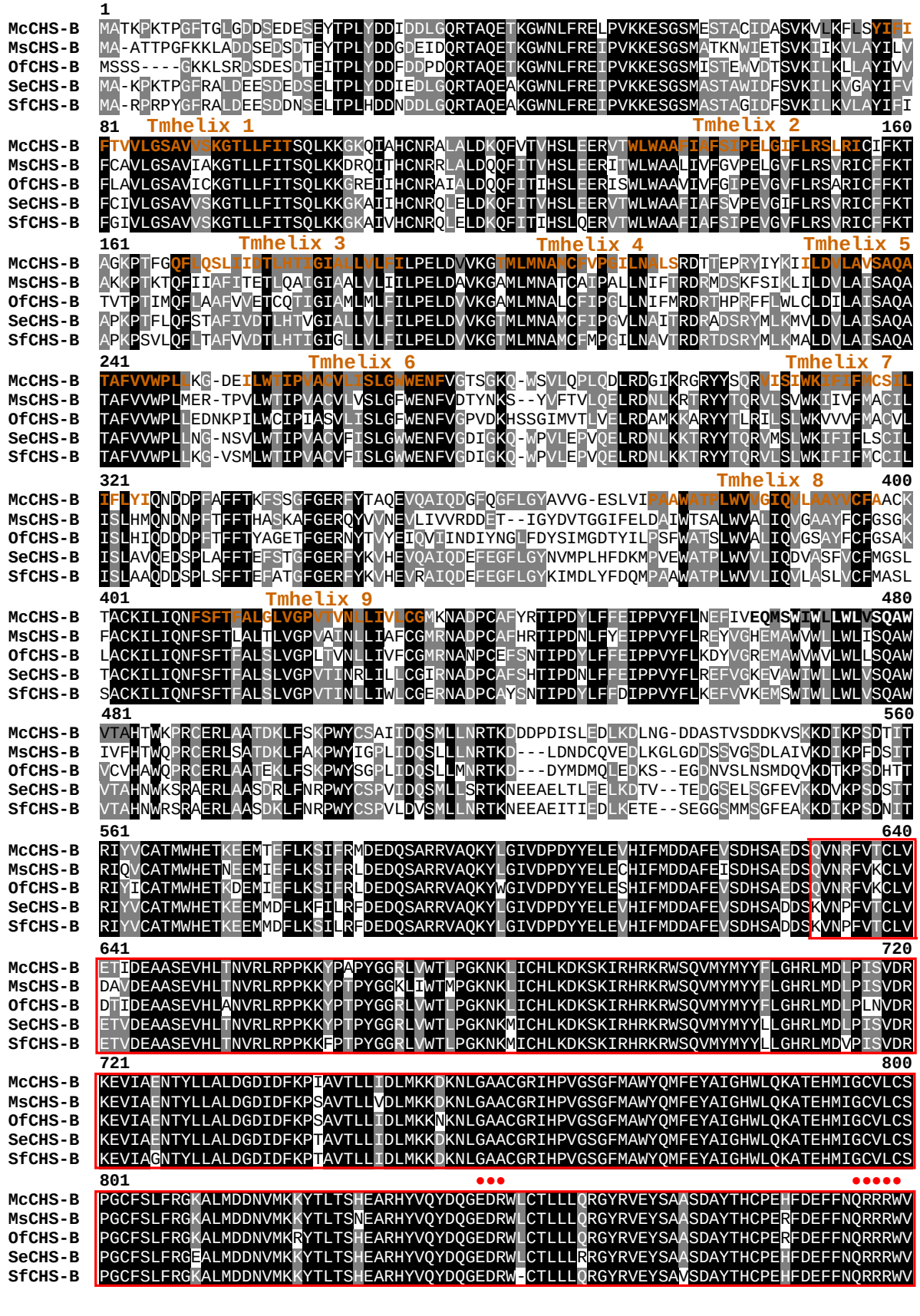
An EST representing a chitinase was not found in the *M. configurata* midgut database; however, RT-PCR with conserved CHI primers was used to amplify a region corresponding to a McCHI mRNA. The 1049 bp fragment encoded a putative protein of 192 amino acids and contained a GH18 chitolectin chitotriosidase domain at the amino terminus and a chitin binding Type 2 domain at the carboxy terminus (Marchler-Bauer et al., 2009). 5' RACE amplified a 1152 bp fragment that corresponded to 370 amino acids and 3' RACE amplified a 660 bp fragment that corresponded to the 3' untranslated region (UTR). The entire 2861 bp cDNA encoded a 562 amino acid protein that was denoted *M. configurata* chitinase (McCHI) (Figure 3.5). This was in accordance with the expected size of the transcript and open reading frame (ORF) of midgut chitinases. McCHI was predicted to have a molecular weight of 60.8 kDa and a signal peptide of 20 amino acids. The protein was predicted to have 22 *O*-glycosylated amino acids corresponding to a linker region (approximately 98 amino acids in length) between the amino terminal GH18

chitolectin chitotriosidase domain (371 amino acids in length) and the carboxy terminal Type 2 chitin binding domain (54 amino acids in length) (Figure 3.5). Two asparagines ($N^{86, 304}$) in the GH18 chitolectin chitotriosidase domain were predicted to be *N*-glycosylated (Figure 3.5). A cysteine register within the CBD consisted of $C^1X_{12}C^2X_5C^3X_9C^4X_{12}C^5X_{10}C^6$.

Alignment of McCHI with other lepidopteran CHIs revealed a high degree of similarity with MbCHI (99.8% identity), *Lacanobia oleracea* CHI (96.3%; LoCHI, CAF05663), *Mythimna separata* CHI (92.3%; MiseCHI, AAS92245), *A. ipsilon* CHI (90.9%; AiCHI, ABW03227), SeCHI (90.9%), SfCHI (89.9%), SlCHI (89.7%), *Hyphantria cunea* CHI (82.9%; HcCHI, AAB47539) and *Ostrinia furnacalis* CHI (78.9%; OfCHI, AAW50396). Sequence identity in the catalytic domain ranged from 88.9 to 100% (positions 26-396 in the amino acid sequence). Additionally, four conserved signature motifs (Watanabe et al., 1993), "KF(M/T)VAVGGW", "FDGLDLDWEYP", "MSYDLRG" and "GA(M/L)TNWAIDMD" were also present (Figure 3.5).

3.3.2.3. N-Acetylglucosaminidase (McNAG)

An EST representing a NAG was not found in the *M. configurata* EST database; however, an RT-PCR reaction using conserved NAG primers produced a single band of approximately 1.6 kb, which was in accordance with the expected size of the region to be amplified. The 1615 bp fragment encoded a protein of 538 amino acids and contained a GH20 *N*-acetylhexosaminidase domain (Marchler-Bauer et al., 2009). The protein was denoted *M. configurata N*-acetylglucosaminidase (McNAG). The partial McNAG was predicted to have a molecular weight of 61.1 kDa. The protein was predicted to have three *N*-glycosylated asparagines at $N^{112, 283, 322}$ with no *O*-glycosylated amino acids and the *N*-acetylhexosaminidase domain is 371 amino acids in length (Figure 3.6). Alignment of McNAG with other lepidopteran NAGs revealed a high degree of similarity with XnNAG (86.5% identity), AiNAG (84.8%), OfNAG (77.8%), BmNAG (75.6%), CfNAG (75.3%), and MsNAG (75.2). The alignment also revealed an incomplete sequence at the amino terminus of McNAG. All McNAG orthologs were predicted to have a signal peptide, suggesting McNAG has also a signal peptide.


McCHS-B
MsCHS-B
OfCHS-B
SeCHS-B
SfCHS-B
Tmhelix 1
Tmhelix 2
Tmhelix 3
Tmhelix 4
Tmhelix 5
Tmhelix 6
Tmhelix 7
Tmhelix 8
Tmhelix 9

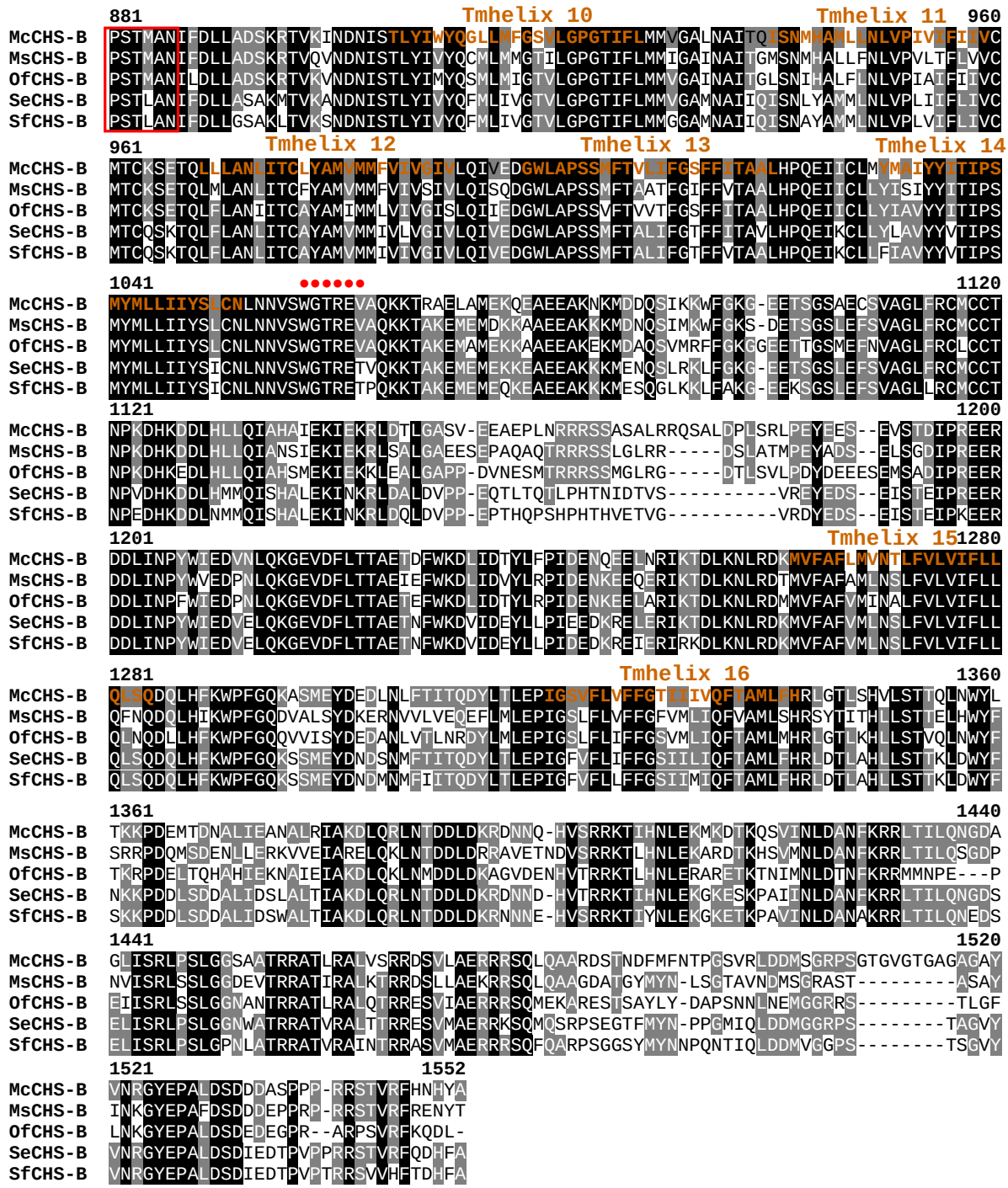
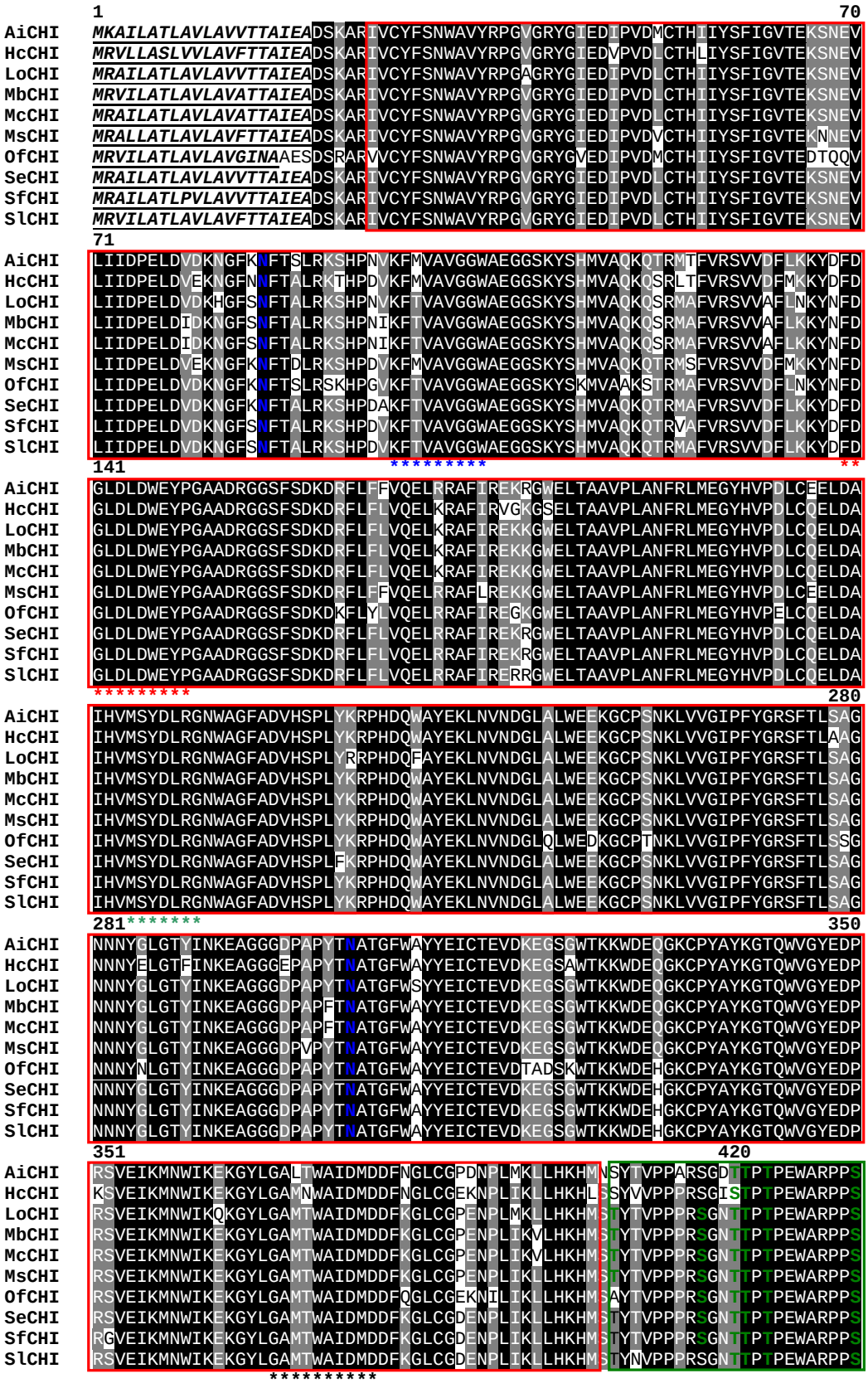


Figure 3.4 Alignment of *Mamestra configurata* chitin synthase-B (McCHS-B) with several lepidopteran chitin synthase-B enzymes from *Manduca sexta* (MsCHS-B, GenBank accesion number, AAX20091), *Spodoptera exigua* (SeCHS-B, ABI96087), *S. frugiperda* (SfCHS-B, AAS12599), and *Ostrinia furnacalis* (OfCHS-B, ABX46067). Chitin synthase domain (in red box), signature motifs “EDR”, “QRRRW”, and “WGTRE” (red •), transmembrane helices (brown lettering) and amino acids present in all (black highlight) or the majority (gray highlight) are shown.


1 70
AiCHI MKAILATLAVLAVVTTAIEADSKARIVCYFSNWAVYRPGVGRYGIEDIPVDMCTHIIYSFIGVTEKSNEV
HcCHI MRVLLASLVVLAVFTTAIEADSKARIVCYFSNWAVYRPGVGRYGIEDVPVDLCTHLIYSFIGVTEKSNEV
LoCHI MRAILATLAVLAVVTTAIEADSKARIVCYFSNWAVYRPGAGRYGIEDIPVDLCTHIIYSFIGVTEKSNEV
MbCHI MRVILATLAVLAVATTAIEADSKARIVCYFSNWAVYRPGVGRYGIEDIPVDLCTHIIYSFIGVTEKSNEV
McCHI MRAILATLAVLAVATTAIEADSKARIVCYFSNWAVYRPGVGRYGIEDIPVDLCTHIIYSFIGVTEKSNEV
MsCHI MRALLATLAVLAVFTTAIEADSKARIVCYFSNWAVYRPGVGRYGIEDIPVDVCTHIIYSFIGVTEKNNEV
OfCHI MRVILATLAVLAVGINAAESDSRARVVCYFSNWAVYRPGVGRYGVEDIPVDMCTHIIYSFIGVTEDTQQV
SeCHI MRAILATLAVLAVVTTAIEADSKARIVCYFSNWAVYRPGVGRYGIEDIPVDLCTHIIYSFIGVTEKSNEV
SfCHI MRAILATLPVLAVVTTAIEADSKARIVCYFSNWAVYRPGVGRYGIEDIPVDLCTHIIYSFIGVTEKSNEV
SlCHI MRVILATLAVLAVFTTAIEADSKARIVCYFSNWAVYRPGVGRYGIEDIPVDLCTHIIYSFIGVTEKSNEV
71
AiCHI LIIDPELDVDKNGFKNFTSLRKSHPNVKFMVAVGGWAEGGSKYSHMVAQKQTRMTFVRSVVDFLKKYDFD
HcCHI LIIDPELDVEKNGFNNFTALRKTHPDVKFMVAVGGWAEGGSKYSHMVAQKQSRLTFVRSVVDFMKKYDFD
LoCHI LIIDPELDVDKHGFSNFTALRKSHPNVKFTVAVGGWAEGGSKYSHMVAQKQSRMAFVRSVVAFLNKYNFD
MbCHI LIIDPELDIDKNGFSNFTALRKSHPNIKFTVAVGGWAEGGSKYSHMVAQKQSRMAFVRSVVAFLKKYNFD
McCHI LIIDPELDIDKNGFSNFTALRKSHPNIKFTVAVGGWAEGGSKYSHMVAQKQSRMAFVRSVVAFLKKYNFD
MsCHI LIIDPELDVEKNGFKNFTDLRKSHPDVKFMVAVGGWAEGGSKYSHMVAQKQTRMSFVRSVVDFMKKYNFD
OfCHI LIIDPELDVDKNGFKNFTSLRSKHPGVKFTVAVGGWAEGGSKYSKMVAAKSTRMAFVRSVVDFLNKYNFD
SeCHI LIIDPELDVDKNGFKNFTALRKSHPDAKFTVAVGGWAEGGSKYSHMVAQKQTRMAFVRSVVDFLKKYDFD
SfCHI LIIDPELDVDKNGFSNFTALRKSHPDVKFTVAVGGWAEGGSKYSHMVAQKQTRVAFVRSVVDFLKKYDFD
SlCHI LIIDPELDVDKNGFSNFTALRKSHPDVKFTVAVGGWAEGGSKYSHMVAQKQTRMAFVRSVVDFLKKYDFD
141 **********
AiCHI GLDLDWEYPGAADRGGSFSDKDRFLFFVQELRRAFIREKRGWELTAAVPLANFRLMEGYHVPDLCEELDA
HcCHI GLDLDWEYPGAADRGGSFSDKDRFLFLVQELKRAFIRVGKGSELTAAVPLANFRLMEGYHVPDLCQELDA
LoCHI GLDLDWEYPGAADRGGSFSDKDRFLFLVQELKRAFIREKKGWELTAAVPLANFRLMEGYHVPDLCQELDA
MbCHI GLDLDWEYPGAADRGGSFSDKDRFLFLVQELKRAFIREKKGWELTAAVPLANFRLMEGYHVPDLCQELDA
McCHI GLDLDWEYPGAADRGGSFSDKDRFLFLVQELKRAFIREKKGWELTAAVPLANFRLMEGYHVPDLCQELDA
MsCHI GLDLDWEYPGAADRGGSFSDKDRFLFFVQELRRAFLREKKGWELTAAVPLANFRLMEGYHVPDLCEELDA
OfCHI GLDLDWEYPGAADRGGSFSDKDKFLYLVQELRRAFIREGKGWELTAAVPLANFRLMEGYHVPELCQELDA
SeCHI GLDLDWEYPGAADRGGSFSDKDRFLFLVQELRRAFIREKRGWELTAAVPLANFRLMEGYHVPDLCQELDA
SfCHI GLDLDWEYPGAADRGGSFSDKDRFLFLVQELRRAFIREKRGWELTAAVPLANFRLMEGYHVPDLCQELDA
SlCHI GLDLDWEYPGAADRGGSFSDKDRFLFLVQELRRAFIRERRGWELTAAVPLANFRLMEGYHVPDLCQELDA
********* 280
AiCHI IHVMSYDLRGNWAGFADVHSPLYKRPHDQWAYEKLNVNDGLALWEEKGCPSNKLVVGIPFYGRSFTLSAG
HcCHI IHVMSYDLRGNWAGFADVHSPLYKRPHDQWAYEKLNVNDGLALWEEKGCPSNKLVVGIPFYGRSFTLAAG
LoCHI IHVMSYDLRGNWAGFADVHSPLYRRPHDQFAYEKLNVNDGLALWEEKGCPSNKLVVGIPFYGRSFTLSAG
MbCHI IHVMSYDLRGNWAGFADVHSPLYKRPHDQWAYEKLNVNDGLALWEEKGCPSNKLVVGIPFYGRSFTLSAG
McCHI IHVMSYDLRGNWAGFADVHSPLYKRPHDQWAYEKLNVNDGLALWEEKGCPSNKLVVGIPFYGRSFTLSAG
MsCHI IHVMSYDLRGNWAGFADVHSPLYKRPHDQWAYEKLNVNDGLALWEEKGCPSNKLVVGIPFYGRSFTLSAG
OfCHI IHVMSYDLRGNWAGFADVHSPLYKRPHDQWAYEKLNVNDGLQLWEDKGCPTNKLVVGIPFYGRSFTLSSG
SeCHI IHVMSYDLRGNWAGFADVHSPLFKRPHDQWAYEKLNVNDGLALWEEKGCPSNKLVVGIPFYGRSFTLSAG
SfCHI IHVMSYDLRGNWAGFADVHSPLYKRPHDQWAYEKLNVNDGLALWEEKGCPSNKLVVGIPFYGRSFTLSAG
SlCHI IHVMSYDLRGNWAGFADVHSPLYKRPHDQWAYEKLNVNDGLALWEEKGCPSNKLVVGIPFYGRSFTLSAG
281******* 350
AiCHI NNNYGLGTYINKEAGGGDPAPYTNATGFWAYYEICTEVDKEGSGWTKKWDEQGKCPYAYKGTQWVGYEDP
HcCHI NNNYELGTFINKEAGGGEPAPYTNATGFWAYYEICTEVDKEGSAWTKKWDEQGKCPYAYKGTQWVGYEDP
LoCHI NNNYGLGTYINKEAGGGDPAPYTNATGFWSYYEICTEVDKEGSGWTKKWDEQGKCPYAYKGTQWVGYEDP
MbCHI NNNYGLGTYINKEAGGGDPAPFTNATGFWAYYEICTEVDKEGSGWTKKWDEQGKCPYAYKGTQWVGYEDP
McCHI NNNYGLGTYINKEAGGGDPAPFTNATGFWAYYEICTEVDKEGSGWTKKWDEQGKCPYAYKGTQWVGYEDP
MsCHI NNNYGLGTYINKEAGGGDPVPYTNATGFWAYYEICTEVDKEGSGWTKKWDEQGKCPYAYKGTQWVGYEDP
OfCHI NNNYNLGTYINKEAGGGDPAPYTNATGFWAYYEICTEVDTADSKWTKKWDEHGKCPYAYKGTQWVGYEDP
SeCHI NNNYGLGTYINKEAGGGDPAPYTNATGFWAYYEICTEVDKEGSGWTKKWDEHGKCPYAYKGTQWVGYEDP
SfCHI NNNYGLGTYINKEAGGGDPAPYTNATGFWAYYEICTEVDKEGSGWTKKWDEHGKCPYAYKGTQWVGYEDP
SlCHI NNNYGLGTYINKEAGGGDPAPYTNATGFWAYYEICTEVDKEGSGWTKKWDEHGKCPYAYKGTQWVGYEDP
351 420
AiCHI RSVEIKMNWIKEKGYLGALTWAIDMDDFNGLCGPDNPLMKLLHKHMNSYTVPPARSGDTTPTPEWARPPS
HcCHI KSVEIKMNWIKEKGYLGAMNWAIDMDDFNGLCGEKNPLIKLLHKHLSSYVVPPPRSGISTPTPEWARPPS
LoCHI RSVEIKMNWIKQKGYLGAMTWAIDMDDFKGLCGPENPLMKLLHKHMSTYTVPPPRSGNTTPTPEWARPPS
MbCHI RSVEIKMNWIKEKGYLGAMTWAIDMDDFKGLCGPENPLIKVLHKHMSTYTVPPPRSGNTTPTPEWARPPS
McCHI RSVEIKMNWIKEKGYLGAMTWAIDMDDFKGLCGPENPLIKVLHKHMSTYTVPPPRSGNTTPTPEWARPPS
MsCHI RSVEIKMNWIKEKGYLGAMTWAIDMDDFKGLCGPENPLIKLLHKHMSTYTVPPPRSGNTTPTPEWARPPS
OfCHI RSVEIKMNWIKEKGYLGAMTWAIDMDDFQGLCGEKNILIKLLHKHMSAYTVPPPRSGNTTPTPEWARPPS
SeCHI RSVEIKMNWIKEKGYLGAMTWAIDMDDFKGLCGDENPLIKLLHKHMSTYTVPPPRSGNTTPTPEWARPPS
SfCHI RGVEIKMNWIKEKGYLGAMTWAIDMDDFKGLCGDENPLIKLLHKHMSTYTVPPPRSGNTTPTPEWARPPS
SlCHI RSVEIKMNWIKEKGYLGAMTWAIDMDDFKGLCGDENPLIKLLHKHMSTYNVPPPRSGNTTPTPEWARPPS
****

(Figure 3.5 continued)

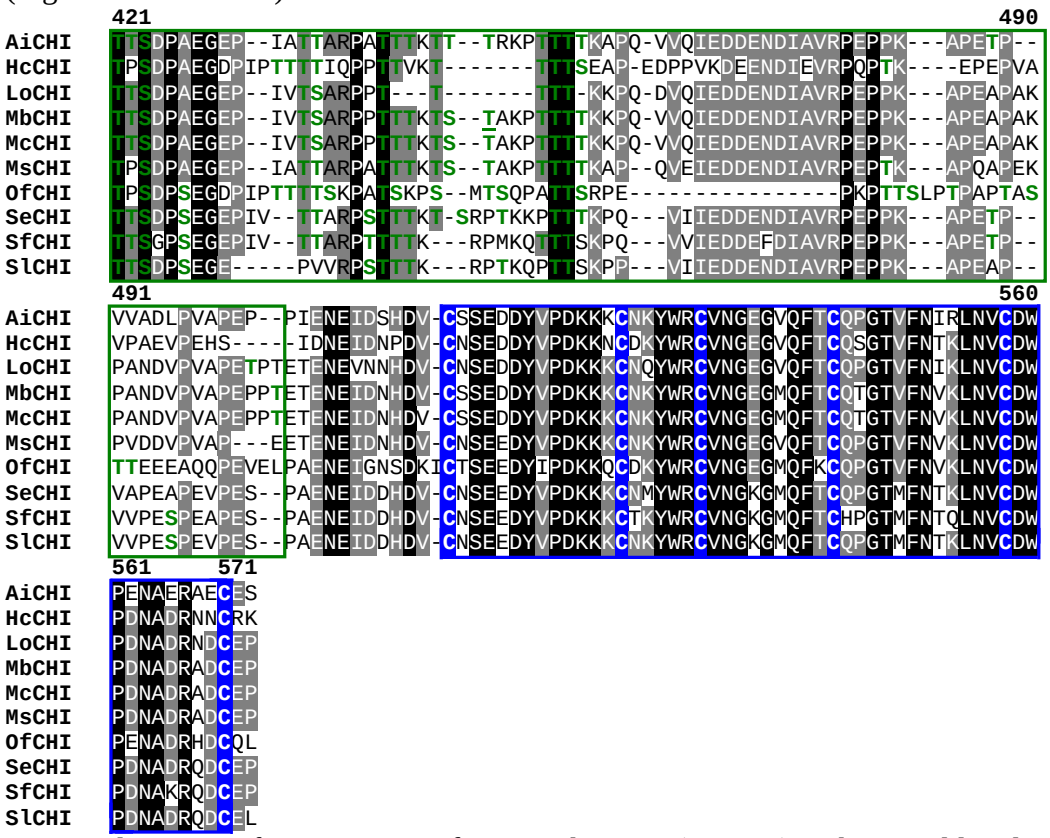
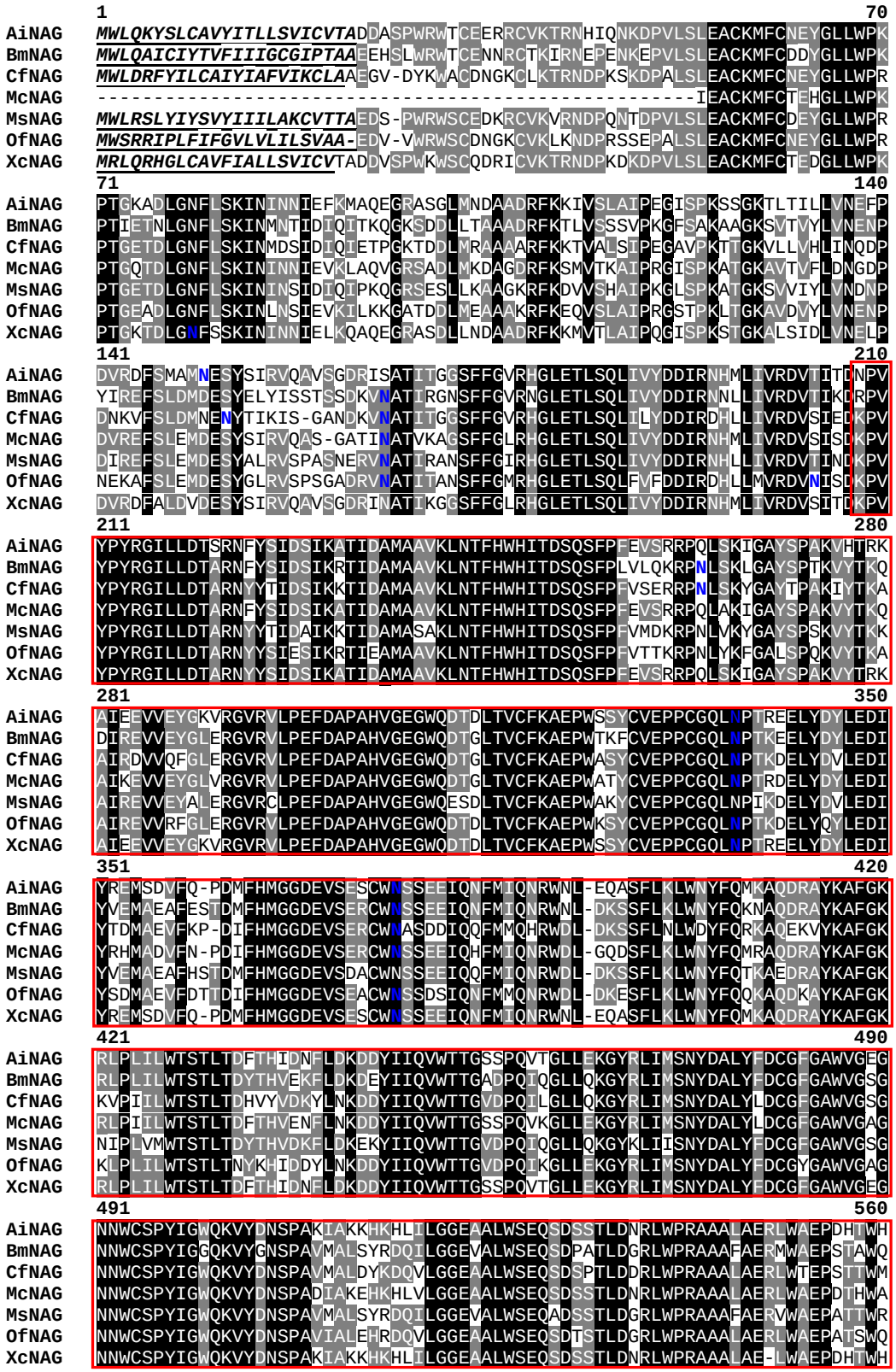


Figure 3.5 Alignment of *Mamestra configurata* chitinase (McCHI) with several lepidopteran chitinases. The enzymes used are from *Agrostis ipsilon* (AiCHI, GenBank accession number: ABW03227), *Hyphantria cunea* (HcCHI, AAB47539), *Lacanobia oleracea* (LoCHI, CAF05663), *Mamestra brassicae* (MbCHI, ACL30984), *Mythimna separata* (MiseCHI, AAS92245), *Ostrinia furnacalis* (OfCHI, AAW50396), *Spodoptera exigua* (SeCHI, ADI24346), *S. frugiperda* (SfCHI, AAS18266) and *S. litura* (SlCHI, BAB12678). Signal peptides (bold, italics and underlined), amino acids present in all (black background) or the majority (gray background), GH18 chitolectin chitotriosidase domain (red box) containing conserved Motif 1 “KXXXXXGGW” (blue *), Motif 2 “FDGXDLDWEYP” (red *), Motif 3 “MXYDXXG” (green *) and Motif 4 "GXXXWXXDXD" (black *) where X could be any amino acid, putative linker region (green box) containing serines and threonines that were predicted to be *O*-glycosylated (green lettering), chitin binding Type 2 domain (blue box) containing conserved cysteines that are involved in disulphide bridge formation (blue background), and asparagines predicted to be *N*-glycosylated (blue lettering) are shown.


1 70
AiNAG MWLQKYSLCAVYITLLSVICVTADDASPWRWTCEERRCVKTRNHIQNKDPVLSLEACKMFCNEYGLLWPK
BmNAG MWLQAICIYTVFIIIGCGIPTAAEEHSLWRWTCENNRCTKIRNEPENKEPVLSLEACKMFCDDYGLLWPK
CfNAG MWLDRFYILCAIYIAFVIKCLAAEGV-DYKWACDNGKCLKTRNDPKSKDPALSLEACKMFCNEYGLLWPR
McNAG IEACKMFCTEHGLLWPK
MsNAG MWLRSLYIYSVYIIILAKCVTTAEDS-PWRWSCEDKRCVKVRNDPQNTDPVLSLEACKMFCDEYGLLWPR
OfNAG MWSRRIPLFIFGVLVLILSVAA-EDV-VWRWSCDNGKCVKLKNDPRSSEPALSLEACKMFCNEYGLLWPR
XcNAG MRLQRHGLCAVFIALLSVICVTADDVSPWKWSCQDRICVKTRNDPKDKDPVLSLEACKMFCTEDGLLWPK
71 140
AiNAG PTGKADLGNFLSKININNIEFKMAQEGRASGLMNDAADRFKKIVSLAIPEGISPKSSGKTLTILLVNEFP
BmNAG PTIETNLGNFLSKINMNTIDIQITKQGKSDDLLTAAADRFKTLVSSVPKGFSAKAAGKSVTVYLVNENP
CfNAG PTGETDLGNFLSKINMDSIDIQIETPGKTDDLMRAAAARFKKTVALSIPEGAVPKTTGKVLLVHLINQDP
McNAG PTGQTDLGNFLSKININNIEVKLAQVGRSADLMKDAGDRFKSMVTKAIPRGISPKATGKAVTVFLDNGDP
MsNAG PTGETDLGNFLSKININSIDIQIPKQGRSESLLKAAGKRFKDVVSHAIPKGLSPKATGKSVVIYLVNDNP
OfNAG PTGEADLGNFLSKINLNSIEVKILKKGATDDLMEAAAKRFKEQVSLAIPRGSTPKLTGKAVDVYLVNENP
XcNAG PTGKTDLGNFSSKININNIELKQAQEGRASDLLNDAADRFKKMVTLAIPQGISPKSTGKALSIDLVNELP
141 210
AiNAG DVRDFSMAMNESYSIRVQAVSGDRISATITGGSFFGVRHGLETLSQLIVYDDIRNHMLIVRDVTITDNPV
BmNAG YIREFSLDMDESYELYISSTSSDKVNATIRGNSFFGVRNGLETLSQLIVYDDIRNNLLIVRDVTIKDRPV
CfNAG DNKVFSLDMNENYTIKIS-GANDKVNATITGGSFFGVRHGLETLSQLILYDDIRDHLLIVRDVSIEDKPV
McNAG DVREFSLEMDESYSIRVQAS-GATINATVKAGSFFGLRHGLETLSQLIVYDDIRNHMLIVRDVSISDKPV
MsNAG DIREFSLEMDESYALRVSPASNERVNATIRANSFFGIRHGLETLSQLIVYDDIRNHLLIVRDVTINDKPV
OfNAG NEKAFSLEMDESYGLRVSPSGADRVNATITANSFFGMRHGLETLSQLFVFDDIRDHLLMVRDVNISDKPV
XcNAG DVRDFALDVDESYSIRVQAVSGDRINATIKGGSFFGLRHGLETLSQLIVYDDIRNHMLIVRDVSITDKPV
211 280
AiNAG YPYRGILLDTSRNFYSIDSIKATIDAMAAVKLNTFHWHITDSQSFPFEVSRRPQLSKIGAYSPAKVHTRK
BmNAG YPYRGILLDTARNFYSIDSIKRTIDAMAAVKLNTFHWHITDSQSFPLVLQKRPNLSKLGAYSPTKVYTKQ
CfNAG YPYRGILLDTARNYYTIDSIKKTIDAMAAVKLNTFHWHITDSQSFPFVSERRPNLSKYGAYTPAKIYTKA
McNAG YPYRGILLDTARNFYSIDSIKATIDAMAAVKLNTFHWHITDSQSFPFEVSRRPQLAKIGAYSPAKVYTKQ
MsNAG YPYRGILLDTARNYYTIDAIKKTIDAMASAKLNTFHWHITDSQSFPFVMDKRPNLVKYGAYSPSKVYTKK
OfNAG YPYRGILLDTARNYYSIESIKRTIEAMAAVKLNTFHWHITDSQSFPFVTTKRPNLYKFGALSPQKVYTKA
XcNAG YPYRGILLDTARNYYSIDSIKATIDAMAAVKLNTFHWHITDSQSFPFEVSRRPQLSKIGAYSPAKVYTRK
281 350
AiNAG AIEEVVEYGKVRGVRVLPEFDAPAHVGEGWQDTDLTVCFKAEPWSSYCVEPPCGQLNPTREELYDYLEDI
BmNAG DIREVVEYGLERGVRVLPEFDAPAHVGEGWQDTGLTVCFKAEPWTKFCVEPPCGQLNPTKEELYDYLEDI
CfNAG AIRDVVQFGLERGVRVLPEFDAPAHVGEGWQDTGLTVCFKAEPWASYCVEPPCGQLNPTKDELYDVLEDI
McNAG AIKEVVEYGLVRGVRVLPEFDAPAHVGEGWQDTGLTVCFKAEPWATYCVEPPCGQLNPTRDELYDYLEDI
MsNAG AIREVVEYALERGVRCLPEFDAPAHVGEGWQESDLTVCFKAEPWAKYCVEPPCGQLNPIKDELYDVLEDI
OfNAG AIREVVRFGLERGVRVLPEFDAPAHVGEGWQDTDLTVCFKAEPWKSYCVEPPCGQLNPTKDELYQYLEDI
XcNAG AIEEVVEYGKVRGVRVLPEFDAPAHVGEGWQDTDLTVCFKAEPWSSYCVEPPCGQLNPTREELYDYLEDI
351 420
AiNAG YREMSDVFQ-PDMFHMGGDEVSESCWNSSEEIQNFMIQNRWNL-EQASFLKLWNYFQMKAQDRAYKAFGK
BmNAG YVEMAEAFESTDMFHMGGDEVSERCWNSSEEIQNFMIQNRWNL-DKSSFLKLWNYFQKNAQDRAYKAFGK
CfNAG YTDMAEVFKP-DIFHMGGDEVSERCWNASDDIQQFMMQHRWDL-DKSSFLNLWDYFQRKAQEKVYKAFGK
McNAG YRHMADVFN-PDIFHMGGDEVSERCWNSSEEIQHFMIQNRWDL-GQDSFLKLWNYFQMRAQDRAYKAFGK
MsNAG YVEMAEAFHSTDMFHMGGDEVSDACWNSSEEIQQFMIQNRWDL-DKSSFLKLWNYFQTKAEDRAYKAFGK
OfNAG YSDMAEVFDTTDIFHMGGDEVSEACWNSSDSIQNFMMQNRWDL-DKESFLKLWNYFQQKAQDKAYKAFGK
XcNAG YREMSDVFQ-PDMFHMGGDEVSESCWNSSEEIQNFMIQNRWNL-EQASFLKLWNYFQMKAQDRAYKAFGK
421 490
AiNAG RLPLILWTSTLTDFTHIDNFLDKDDYIIQVWTTGSSPQVTGLLEKGYRLIMSNYDALYFDCGFGAWVGEG
BmNAG RLPLILWTSTLTDYTHVEKFLDKDEYIIQVWTTGADPQIQGLLQKGYRLIMSNYDALYFDCGFGAWVGSG
CfNAG KVPIILWTSTLTDHVYVDKYLNKDDYIIQVWTTGVDPQILGLLQKGYRLIMSNYDALYLDCGFGAWVGSG
McNAG RLPIILWTSTLTDFTHVENFLNKDDYIIQVWTTGSSPQVKGLLEKGYRLIMSNYDALYLDCGFGAWVGAG
MsNAG NIPLVMWTSTLTDYTHVDKFLDKEKYIIQVWTTGVDPQIQGLLQKGYKLIISNYDALYFDCGFGAWVGSG
OfNAG KLPLILWTSTLTNYKHIDDYLNKDDYIIQVWTTGVDPQIKGLLEKGYRLIMSNYDALYFDCGYGAWVGAG
XcNAG RLPLILWTSTLTDFTHIDNFLDKDDYIIQVWTTGSSPQVTGLLEKGYRLIMSNYDALYFDCGFGAWVGEG
491 560
AiNAG NNWCSPYIGWQKVYDNSPAKIAKKHKHLILGGEAALWSEQSDSSTLDNRLWPRAAALAERLWAEPDHTWH
BmNAG NNWCSPYIGGQKVYGNSPAVMALSYRDQILGGEVALWSEQSDPATLDGRLWPRAAAFAERMWAEPSTAWQ
CfNAG NNWCSPYIGWQKVYDNSPAVMALDYKDQVLGGEAALWSEQSDSPTLDDRLWPRAAALAERLWTEPSTTWM
McNAG NNWCSPYIGWQKVYDNSPADIAKEHKHLVLGGEAALWSEQSDSSTLDNRLWPRAAALAERLWAEPDTHWA
MsNAG NNWCSPYIGWQKVYDNSPAVMALSYRDQILGGEVALWSEQADSSTLDGRLWPRAAAFAERVWAEPATTWR
OfNAG NNWCSPYIGWQKVYDNSPAVIALEHRDQVLGGEAALWSEQSDTSTLDGRLWPRAAALAERLWAEPATSWQ
XcNAG NNWCSPYIGWQKVYDNSPAKIAKKHKHLILGGEAALWSEQSDSSTLDNRLWPRAAALAE-LWAEPDHTWH

(Figure 3.6 continued)

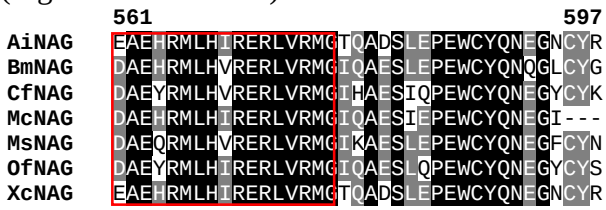


Figure 3.6 Alignment of *Mamestra configurata N*-acetylglucosaminidase (McNAG) with several lepidopteran *N*-acetylglucosaminidases. The enzymes used were from *Agrotis ipsilon* (AiNAG, GU985280), *Bombyx mori* (BmNAG, NM_001044002), *Choristoneura fumiferana* (CfNAG, DQ005717), *Manduca sexta* (MsNAG, AY368703), *Ostrinia furnacalis* (OfNAG, DQ887769), and *Xestia c-nigrum* (XnNAG, FJ848784). Signal peptides (bold, italics and underlined), asparagines predicted to be *N*-glycosylated (blue lettering), GH20 *N*-acetylhexosaminidase domain (in red box) and amino acids present in all (black background) or the majority (gray background) are shown.

3.3.3 Identification of M. configurata *Ortholog of Putative Chitin Synthase Activator, MsCTLP1*

BLAST analysis of the *M. configurata* midgut EST database with *M. sexta* MsCTLP1 revealed that McSP33 (Erlandson et al., 2010) was the MsCTLP1 ortholog in *M. configurata*. The two proteins shared 63% identity. The 974 bp McSP33 cDNA was predicted to encode a chymotrypsin as it had a glycine at position 230, which corresponds to specificity amino acid (G^{189}) of bovine chymotrypsin (Figure 3.7). *McSP33* has an ORF of 279 amino acids with a predicted molecular weight of 27.5 kDa after removal of an 18 amino acid signal peptide (Figure 3.7). The catalytic triad of both McSP33 and MsCTLP1 contains H^{57}, D^{102}, and S^{195}. A conserved trypsin cleavage site, "RI" was also detected in both proteins and each had six conserved cysteines.

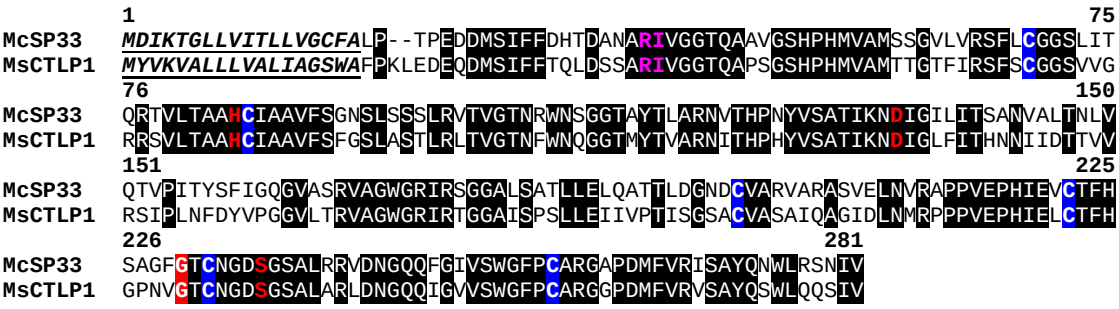


Figure 3.7 Alignment of *Mamestra configurata* McSP33 with *Manduca sexta* MsCTLP1. The signal peptide (bold, underlined and italics), putative trypsin cleavage site (pink amino acids), catalytic triad containing the conserved H^{57}, D^{102}, and S^{195} residues (red lettering), specificity amino acid (red background), conserved cysteine (blue background), amino acids present in both black background)are shown. GenBank accession numbers for McSP33 and MsCTLP1 are FJ205415 and AM419170, respectively.

3.3.4 Expression Analysis of the Genes Encoding Midgut Chitin Synthase and Chitinolytic Enzymes

RT-PCR analysis indicated that *McCHS-B* was expressed only in midgut. The *McCHI* and *McNAG* were expressed in the foregut, midgut, hindgut, Malpighian tubules, integument, and tracheae with differences in the level of expression (Figure 3.8).

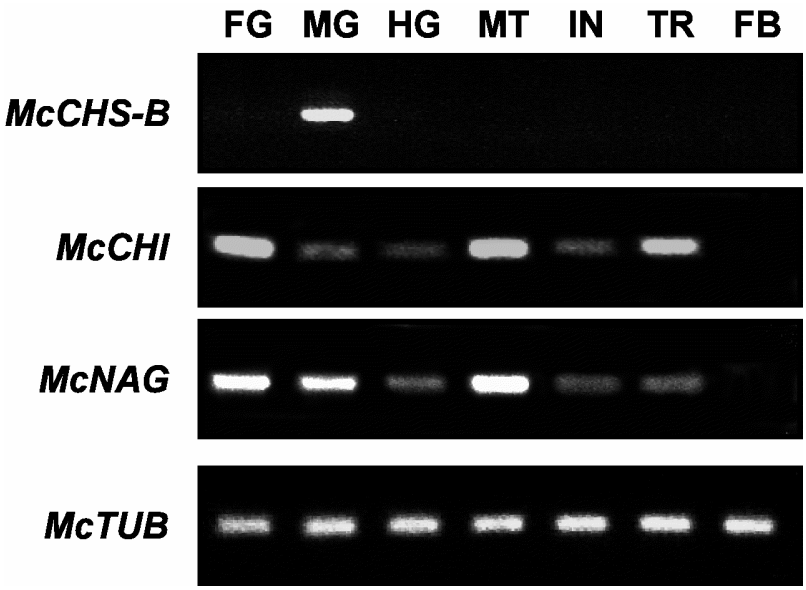


Figure 3.8 Tissue specific expression of *Mamestra configurata* chitin synthase-B (*McCHS-B*), chitinase (*McCHI*) and *N*-acetylglucosaminidase (*McNAG*). Gene expression patterns were determined using total RNA from foregut (FG), midgut (MG), hindgut (HG), Malpighian tubules (MT), integument (IN), tracheae (TR) and fat body (FB) from 5th instar larvae by RT-PCR analysis. Amplification of the *M. configurata tubulin* (*McTUB)* gene was used as a control in the RT-PCR.

qPCR was conducted to quantify the changes in *McCHS-B* and *McCHI* expression during larval development. This analysis showed that *McCHI* was strongly upregulated at the molting stage; approximately 15, 12, 19 times more in comparison to that in middle stage feeding for Molting Phases I, II, and III, respectively, but its expression was low at feeding stages (Figure 3.9). Although *McCHI* expression was lower in starved larvae than in molting larvae, it was still approximately 3.5 times higher than in middle stage feeding larvae (Figure 3.9). *McCHS-B* expression was found to be fairly consistent in all samples (Figure 3.9)

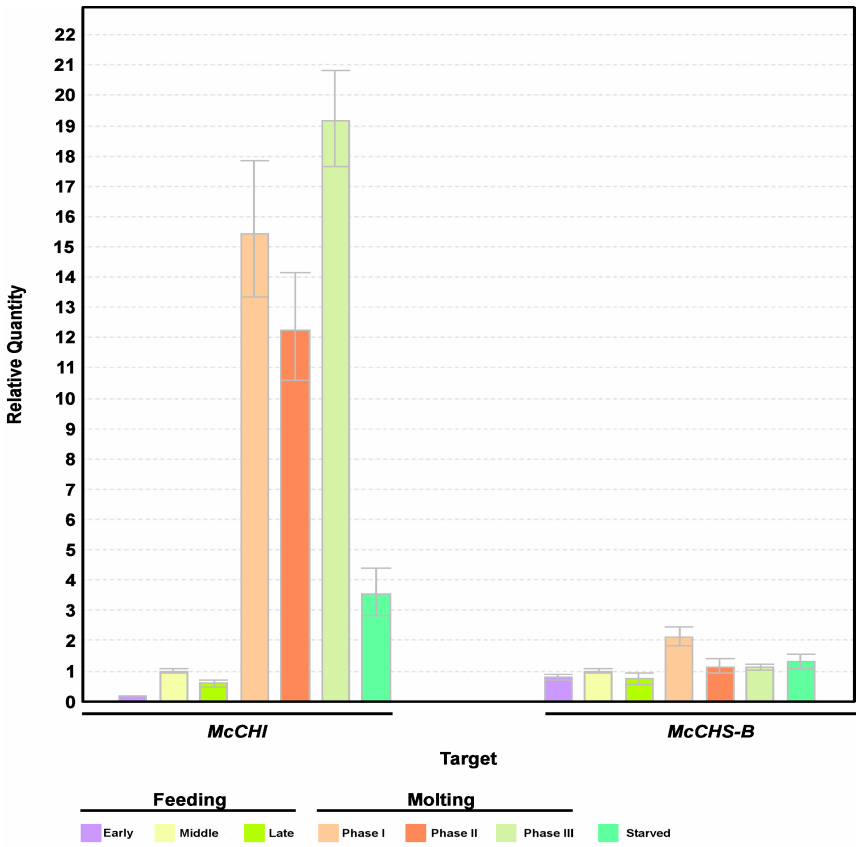


Figure 3.9 Expression of *Mamestra configurata* midgut chitinase (*CHI*) and chitin synthase-B (*CHS-B*) as determined by qPCR in different developmental stages. Total RNA from the midgut of early, middle, late stage feeding and 24 h starved 5th instar larvae and Phase I, Phase II, and Phase III larvae molting from 5th to 6th instar was used for qPCR analysis. Amplification of the *M. configurata* actin (*McACT*) gene was used as the endogenous reference gene in the qPCR. All values are relative to the middle feeding stage which is represented as 1.

RT-PCR analysis revealed that *McNAG* was expressed at low levels in the early and middle stage feeding larvae, with an increase starting by late feeding, followed by a strong upregulation in Phase I and II molting stages. In contrast, there was a decrease in its expression during molting Phase III stage and starvation, though the decrease in the latter was dramatic (Figure 3.10). Like *McCHS-B*, expression of *McSP33* was consistent throughout larval development (Figure 3.10).

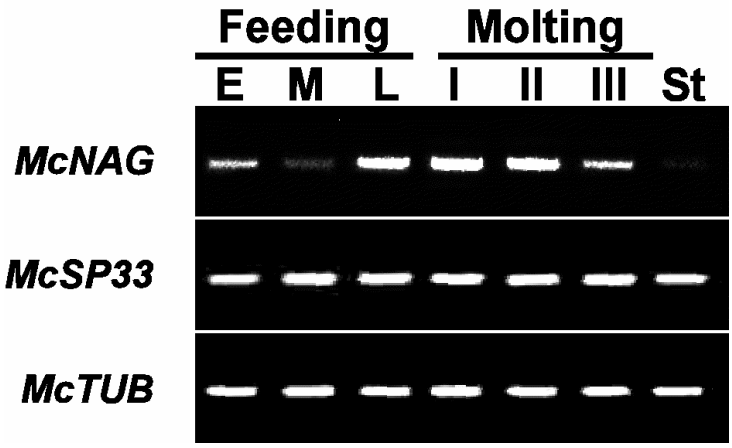


Figure 3.10 Expression of *Mamestra configurata N*-acetylglucosaminidase (*McNAG*) and serine protease 33 (*McSP33*) genes. Total RNA from the midgut of early (E), middle (M) and late (L) stage 5th instar feeding, and Phases I, II and III molting (from 5th to 6th instar) larvae was used for RT-PCR analysis. Amplification of the *M. configurata* tubulin (*McTUB*) gene was used as a control in the RT-PCR.

3.3.5 *β-N-Acetylglucosaminidase Activity in PMs from Feeding, Starved and Molting Larvae*

NAG activity within the PM peaked in Phase I molting larvae, while the activity in all other stages, including feeding, molting Phase II, Phase III and starved larvae, was significantly lower (Figure 3.11).

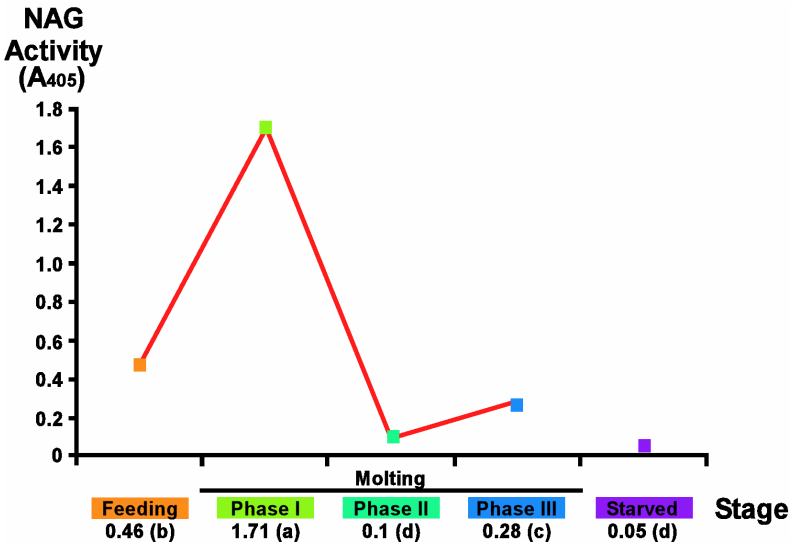


Figure 3.11 β-*N*-Acetylglucosaminidase (NAG) activity in the PM of 5th instar feeding, Phase I, Phase II, and Phase III molting (from 5th to 6th) and 5th instar larvae starved for 24 h. The activity is represented as the amount of chromophore released when measured at A_{405} nm. Letters in parentheses indicate statistically significant differences (one way ANOVA, P=0.05; Tukey HSD, P=0.05).

3.4 Discussion

3.4.1 Physical Architecture of the M. configurata *PM during Development*

Most lepidopteran larvae and adult female mosquitoes produce a Type I PM, though in the latter, it is produced only in response to a blood meal (Freyvogel and Jaquet, 1965) while the *M. configurata* larval PM is also present in nonfeeding stages with the possible exception of early molting larvae. Lack of a PM in early molting larvae was also reported from other lepidopterans, such as *T. ni* (Harper and Granados, 1999) and *Helicoverpa armigera* (Campbell et al., 2008). In feeding and starved *M. configurata* larvae, the PM was found to have organized striations, possibly chitin bundles, running perpendicular to the long axis of the PM. This may be due to periodic pulses of PM formation which lead to a network moulded upon the epithelial cells that secrete its components (Peters et al., 1979; Harper and Hopkins, 1997). In other insects, the PM may appear as a lattice-like structure organized into an orthogonal or hexagonal network [e.g., *Ostrinia nubilalis* (Harper and Hopkins, 1997), *Diatraea grandiosella* (Daves et al., 2007), *Tineola bisselliella* (Lagermalm et al., 1950)], as a random felt-like collection of fibers [e.g., *T. ni* (Adang and Spence, 1981) and *M. sexta* (Hopkins and Harper, 2001)] or as a smooth felt-like surface without a hexagonal or orthogonal matrix [e.g., *Heliothis virescens* (Ryerse et al., 1992)].

Epifluorescence microscopy using FB28 revealed that the PM of feeding, starved and molting larvae was composed of at least two layers. Confocal microscopy with PMs stained with AWGA provided further details for PMs from feeding larvae. A thicker inner layer was found to surround the bolus, while an outer layer surrounded the entire PM. In addition, 2-10 thinner layers were visible between the two main inner and outer layers in the PMs from 4^{th} and 5^{th} instar feeding larvae, suggesting that pulses of PM formation occur in feeding larvae which delaminate from the midgut epithelium. This is somewhat analogous to the situation in Type I PM synthesis in mosquitoes, where a new PM is formed after each blood meal which then surrounds the PM that was synthesized during the previous meal (Waterhouse, 1953a; b). Therefore, depending on the time of extraction, *M. configurata* larvae may have PMs with varying numbers of layers. The multilayered structure is a common feature of the Type I PM and has been shown in other lepidopteran larvae (Lagermalm et al., 1950; Mercer and Day, 1952; Adang and Spence, 1981) and dipteran adults (Berner et al., 1983; Chapman, 1985; Nagel and Peters, 1991). In contrast, Type II PMs generally lack multiple layers, though they may comprise several strands or laminae (Waterhouse, 1954; Ryerse et al., 1992). PMs of *Orgyia pseudotsugata* (Brandt et al., 1978), and

H. virescens (Ryerse et al., 1992) larvae were shown to have two primary laminae, one facing the endoperitrophic and the other facing the ectoperitrophic space, suggesting PM structure and formation vary among lepidopterans.

One of the interesting findings in the current study is the unique architecture detected in Phase II molting larvae where an opaque inner layer is surrounded by a translucent sleeve. The outer layer may be newly synthesized PM as it was nonstriated, and stained less intensely with FB28, suggesting a lower chitin content. The inner material may be "maturing PM", a situation analogous to the maturation of cuticle following the molt in which secreted cuticle components undergo physical and chemical rearrangements as the cuticle hardens (Gillott, 2005).

Interestingly, SEM analysis of PMs from *M. configurata* and *M. sexta* in this study, as well as reports from *A. ipsilon* (Schmidt et al., 2009), *H. virescens* (Ryerse et al., 1992; Plymale et al., 2008), *O. pseudotsugata* (Brandt et al., 1978) and *T. ni* (Adang and Spence, 1981), revealed the absence of a mesh-like network of chitin fibrils. Bolognesi et al. (2005) used fluorescein isothiocyanate in conjunction with epifluorescence microscopy to examine the PM of *S. frugiperda* during development; however, no structural details were revealed. In my study, treatment of the PMs with FB28 or AWGA in conjunction with epifluorescence or confocal microscopy, respectively, revealed details of the PM architecture, indicating that these are good tools for PM structural analysis.

3.4.2 Features of McCHS-B, McCHI, McNAG and McSP33

McCHS-B was predicted to lack a signal peptide, but has nine and seven membrane spanning regions before and after the catalytic domain, respectively, suggesting that it is a membrane bound protein. CHS-Bs have been reported to be membrane bound proteins with membrane spanning regions flanking the catalytic domain. The catalytic domain is located on the cytoplasmic side of the membrane and catalyzes the incorporation of GlcNAc from UDP-GlcNAc into chitin (Tellam et al., 2000; Merzendorfer, 2006). McCHS-B has the signature motifs ("EDR", "QRRRW" and "WGTRE") which are essential for enzyme activity (Hogenkamp et al., 2005).

McCHI consists of an amino terminal catalytic domain and a carboxy terminal cysteine-rich chitin binding domain that are separated by a serine/threonine/proline-rich linker region. A signal peptide was also predicted, suggesting that the protein is secreted. The 18 glycosyl

hydrolase catalytic domain is generally 370 amino acids in length and has four conserved signature motifs (Watanabe et al., 1993; Van Scheltinga et al., 1994; Kramer and Muthukrishnan, 2005) required for enzyme activity (Huang et al., 2000; Lu et al., 2002; Royer et al., 2002). McCHI, as well as other lepidopteran chitinases, have these motifs. The linker region is approximately 100 amino acids in length and predicted to be highly *O*-glycosylated (Arakane et al., 2003) which apparently facilitates secretion of the enzyme from the cell (Zhu et al., 2001). McCHI was predicted to have 22 *O*-glycosylated amino acids in this region. The Type 2 CBD at the carboxy terminus of McCHI is a typical Peritrophin-A domain comprising approximately 60 amino acids and is found in most midgut chitinases. This domain was shown to facilitate hydrolytic processes by anchoring the enzyme to the chitin (Arakane et al., 2003), but is not essential for enzyme activity (Zhu et al., 2001). It is noteworthy that the "amino terminal catalytic domain-linker region-CBD" structural organization was also reported for *Tribolium castaneum* chitinases expressed in the integument (Zhu et al., 2008a). This structural organization is common in several other chitinases expressed in the midgut of lepidopterans (Ahmad et al., 2003; Zheng et al., 2003; Bolognesi et al., 2005) and dipterans (Shen and Jacobs-Lorena, 1997; Ramalho-Ortigao and Traub-Cseko, 2003; Dinglasan et al., 2009).

Lepidopteran orthologs of McNAGs all have a signal peptide, suggesting McNAG is also secreted. McNAG as well as its lepidopteran orthologs have from one to five putative *N*-glycosylated asparagines but no *O*-glycosylated amino acids. The *N*-acetylhexosaminidase domain of lepidopteran NAGs is composed of approximately 370 amino acids, which is similar in size to catalytic domains in endochitinases; however, the overall structural organization is different between NAGs and CHIs in lepidopterans, as NAGs do not have a CBD or linker region rich in serines or threonines.

McSP33 and its ortholog MsCTLP1 have features typical of digestive serine proteases, including the catalytic triad H^{57}, D^{102} and S^{195}. The presence of a glycine at position 189 in both serine proteases suggests they have chymotrypsin-like activity. Furthermore, phylogenetic analysis revealed that McSP33 clusters with other chymotrypsins (Erlandson et al., 2010). A conserved trypsin cleavage site, "RI" was also detected in both proteins, suggesting they are secreted as inactive precursors and activated by trypsin proteolysis (Broehan et al., 2007).

3.4.3 Tissue Specific Expression of the Genes Encoding Midgut Chitin Synthase and Chitinolytic Enzymes

In the current study, *McCHS-B* expression was found to be specific to the midgut, suggesting that it is responsible for the synthesis of chitin in the PM. This finding was also reported for other *CHS-B* genes from lepidopterans (Zimoch and Merzendorfer, 2002; Bolognesi et al., 2005; Kumar et al., 2008), dipterans (Ibrahim et al., 2000; Gagou et al., 2002) and coleopterans (Arakane et al., 2004). In addition, downregulation of *CHS-B* by RNAi in *T. castaneum* reduced chitin content in the PM; further indicating that CHS-B catalyzes the synthesis of chitin in PM (Arakane et al., 2005). Expression of *McCHI* and *McNAG* genes was detected in the foregut, midgut, hindgut, Malpighian tubules, and integument, suggesting that chitinases may be involved in the hydrolysis of chitin in multiple tissues. Another lepidopteran chitinase gene, *H. armigera* chitinase, was also expressed in the midgut, integument and fat body (Ahmad et al., 2003), while *C. fumiferana* chitinase (*CfCHI*) gene was expressed only in the integument and fat body (Zheng et al., 2003). Several other studies revealed chitinase genes that were expressed primarily in midgut (Girard and Jouanin, 1999; Ramalho-Ortigao and Traub-Cseko, 2003; Genta et al., 2006) or midgut and epidermis (Kramer et al., 1993; Kim et al., 1998; Bolognesi et al., 2005). Zhu et al. (2008b) classified the chitinases based on their gene expression patterns, localization, structural organization, and phylogenetic relationship in *T. castaneum*, *Drosophila melanogaster*, and *Anopheles gambiae*. In *T. castaneum*, chitinases expressed in the integument were classified as Group I, while chitinases expressed in the gut were classified as Group IV (Zhu et al., 2008b). As mentioned above, such a classification scheme may not be broadly applicable, since genes encoding chinases with similar structural organization may be expressed in other tissues in species from different orders.

Expression of chitinase genes in the midgut where the PM occurs, the presence of a CBD in the enzyme, and the fact that the PM contains chitin, suggest that they may be physically associated with the PM. Such an association was demonstrated for Group I chitinases in *C. fumiferana* (Zheng et al., 2003) and *A. gambiae* (Dinglasan et al., 2009), suggesting that members of this group may be involved in PM turnover.

3.4.4 Chitin Metabolism in PM during Development

In *M. configurata*, the levels of *CHS-B* expression are constant, with only a slightly higher level in molting Phase I larvae. Consistent *CHS-B* expression agrees with the observation that the Type I PM of lepidopterans is continuously present. In contrast, the Type I PM of mosquitoes is formed in response to feeding, in parallel with the upregulation of *CHS-B* expression (Ibrahim et al., 2000).

McCHI and *McNAG* expression is more dynamic among the various developmental stages than that of *McCHS-B. McCHI and McNAG* are dramatically upregulated during the molt, clearly suggesting the PM is degraded by the simultaneous action of McCHI and McNAG during the molt. This may account for the absence of a PM or the smaller PM in molting Phase I larvae. Furthermore, although endochitinase activity was not measured in the current study, the peak of NAG exochitinase activity in molting Phase I larvae is in accordance with the upregulation of *McCHI*. Thus, the PM from the previous larval instar is degraded and likely to be extruded towards posterior midgut by molting Phase I.

By molting Phase II, a new PM, which is composed of an inner, opaque and an outer, translucent layer, is present. Thus in this way the larva is prepared for the imminent onset of feeding, which starts approximately 30 min-1 h after the Phase II molting stage. The inner PM is amorphous and reacts strongly with the chitin specific FB28, while the outer PM is translucent due to weaker interaction with FB28. These observations suggest that the inner PM is relatively mature or in the process of maturation, while the outer PM is newly secreted and has a lower chitin content.

McCHS-B is expressed continuously at similar levels in all stages. Therefore, its activity may not be controlled at the level of transcription, but rather by activation of the proenzyme. Trypsin has been implicated in the stimulation of chitin synthesis (Cohen and Casida, 1980; Mayer et al., 1980; Merz et al., 1999); however, trypsin does not directly act on the chitin synthase, but on a soluble activating factor (Zimoch et al., 2005). This factor may be the chymotrypsin, CTLP1, that was proposed to activate CHS-B by cleaving near the extracellular carboxy terminal domain (Broehan et al., 2007). However, the *M. configurata* ortholog encoding the putative CHS-B activator, McSP33, is also expressed at a constant level in all situations tested, therefore, there must be another mechanism to control either CHS-B or activator activity. A possible candidate may be the *M. configurata* midgut serpin isoform

(Chamankhah et al., 2003), McSerpin1A, that could negatively regulate the activity of the McSP33. This hypothesis is supported by the fact that *McSerpin1* is downregulated in Phase II molting larvae (Chamankhah et al., 2003) when the new PM is forming. Similarly, upregulation of *McSerpin1* in starved larvae (Chamankhah et al., 2003) would prevent CHS-B activation by the McSP33 and limit renewal of the PM.

Dramatic upregulation of McCHI and McNAG genes in Phase I molting stage coupled with the increased level of NAG activity suggests that chitin degradation is controlled at the level of transcription. Although both genes are also expressed strongly in molting Phase II stage, NAG activity dramatically decreased. This suggests that an additional control mechanism is likely to occur for chitinolytic enzymes. Indeed, biochemical studies on CHI revealed that this enzyme is also zymogenic, requiring proteolytic processing by serine proteases for activation (Koga et al., 1989). This additional mechanism is necessary for the protection of the newly formed PM in molting phase II to avoid degradation by chitinolytic activity. Thus, chitinolytic activity is controlled both at transcriptional and posttranslational levels.

Although *McCHI* and *McNAG* are expressed mainly in the molting stages, they are also expressed at low levels in feeding larvae. Indeed, McNAG activity is higher in feeding larvae than in molting Phase II or III larvae. Expression of chitinase genes during feeding suggests that these enzymes may also play a role in subtle trimming or modifying the PM, other than in simple degradation. A basal level of chitinase expression during feeding was also detected in *H. armigera* larvae (Ahmad et al., 2003) and beetles (Girard and Jouanin, 1999; Zhu et al., 2008b). In mosquitoes, increases in chitinase gene expression or activity were observed upon feeding (Shen and Jacobs-Lorena, 1997; Filho et al., 2002; Ramalho-Ortigao and Traub-Cseko, 2003). Use of a chitinase inhibitor led to the formation of an atypical, thicker PM, confirming that a basal level of chitinase activity is required during feeding for proper PM formation (Filho et al., 2002).

Interestingly, *McCHI* is upregulated in starved larvae, which may be the reason for the deformation in the PMs of starved larvae observed by epifluorescence microscopy. Indeed, Kramer and Muthukrishnan (1997) reported that high levels of chitinases may damage the PM and it may be that during periods of starvation the PM is turned over or the rate of its production slowed. Lehane (1997) also reported that many insects interrupt PM production during starvation and Type I PM formation was lower in starved larvae in comparison to feeding larvae (Aubertot,

1932; Waterhouse, 1954). In accordance with these observations, Zimoch et al. (2005) reported that *MsCHS-B* was downregulated during starvation along with the cessation of PM formation; however, *McCHS-B* was not downregulated in this circumstance.

In conclusion, a multilayered PM is present in feeding and non-feeding stages (except at the end of the Phase I molting stage). PM synthesis and degradation are under the control of McCHS-B, McCHI and McNAG. *McCHS-B* is expressed at similar levels throughout development, while *McCHI* and *McNAG* are upregulated during the molt resulting in degradation of the PM in molting Phase I stage, as evident by the high level of NAG activity. A new PM is formed during the molting Phase II stage. Although *McCHI* and *McNAG* are expressed at high levels in molting Phase II stage, NAG activity is dramatically reduced. *McSP33*, the gene encoding putative CHS-B activator, is expressed continuously similar to *McCHS-B* expression. These findings contradict the commonly held belief that PM synthesis and degradation is controlled by the balance between chitin synthase and chitinase at the level of transcription (Bolognesi et al., 2005). Interestingly, *McCHI* and *McNAG* are expressed in feeding larvae, indicating they may also be involved in the modulation of PM structure rather than only degrading it during the molt. Starvation induces *McCHI* gene expression and may cause the PM to be disrupted.

In the following chapter, the other major PM component, protein, will be the focus. The comprehensive suite of *M. configurata* PM proteins is delineated using proteomic and genomic approaches. In addition, expression of the genes encoding PM proteins along with their structural features is examined. These data are used to predict the putative roles of these proteins in the PM.

4. SURVEY AND PRELIMINARY CHARACTERIZATION OF PROTEINS ASSOCIATED WITH THE PERITROPHIC MATRIX[2]

4.1 Introduction

Due to the varied and essential roles the peritrophic matrix (PM) plays in insect digestive physiology, many studies have focused on PM proteins and their contribution to PM structure and function. Early studies examined the PM proteins from adult dipterans (Stamm et al., 1978; Dorner and Peters, 1988; Ramos et al., 1994). Digestive enzymes were the first proteins reported to be associated with the PM (Terra et al., 1979; Walker et al., 1980). The first PM structural protein characterized was a small chitin binding glycoprotein from sheep blowfly, *Lucilia cuprina*, larvae (Elvin et al., 1996). Subsequently, midgut genomic (Lehane et al., 2003; Shi et al., 2004; Ramalho-Ortigao et al., 2007; Simpson et al., 2007; Jochim et al., 2008; Morris et al., 2009; Pauchet et al., 2009b; Venancio et al., 2009; Pauchet et al., 2010) and proteomic (Ferreira et al., 2007; Campbell et al., 2008; Pauchet et al., 2008; Dinglasan et al., 2009; Morris et al., 2009) analyses revealed a much larger suite of proteins definitively or putatively associated with the PM of coleopterans, dipterans and lepidopterans. To date, little is known about PM proteins from hymenopterans (Marques-Silva et al., 2005).

PM proteins can be divided into enzymes (nonstructural proteins) and peritrophins (structural proteins). PM associated enzymes are involved in digestion of food (e.g., trypsins, chymotrypsins, aminopeptidases, carboxypeptidases, lipases), chitin modification (e.g., chitin deacetylases) or other enzymatic reactions (e.g., alkaline phosphatases) (Terra et al., 1979; Guo et al., 2005; Campbell et al., 2008). Peritrophins are integral PM proteins and in part determine PM structural features such as strength, elasticity and porosity (Shi et al., 2004; Wang et al., 2004a). A feature common to peritrophins is the presence of cysteine rich regions that form chitin binding domains (CBDs) (Tellam et al., 1999; Wang et al., 2004a). Peritrophin CBDs were subdivided into Peritrophin A domain (PAD), Peritrophin B domain (PBD), and Peritrophin C domain (PCD) with registers of 6, 8 and 10 cysteines, respectively (Tellam et al., 1999); however, PAD is the most common. Peritrophins may have mucin domains (MDs) that are rich in threonines, serines and prolines with a high potential for *O*-linked glycosylation (Devine and McKenzie, 1992). Peritrophins with both domain types are referred to as insect intestinal mucins

[2] Part of this chapter is published in *Insect Molecular Biology*, **17:** 573-585, 2008.

(IIMs) (Wang and Granados, 1997b), while those with only CBDs are described as nonmucin peritrophins (Shi et al., 2004).

Previously, only an IIM (McIIM1) and a nonmucin peritrophin (McPM1) were reported from *Mamestra configurata* PM (Shi et al., 2004). In the current chapter, a comprehensive suite of *M. configurata* PM proteins is elucidated with a combination of genomic and proteomic analyses. The structural characteristics and putative functions of PM proteins are presented along with their tissue specific localization in terms of their transcriptional expression.

4.2 Material and Methods

4.2.1 Protein Extraction, Separation and Liquid Chromatography Tandem-Mass Spectrometry (LC-MS/MS)

The strongly associated protein fraction was extracted from 20 PMs dissected from 5th instar larvae and were separated using one dimensional (1D) or two dimensional (2D) gel electrophoresis as described in Chapter 2. Coomassie blue stained slices excised from 1D SDS-PAGE gel were subjected to LC-MS/MS as described in Chapter 2.

4.2.2 Expression of the Genes Encoding PM Proteins

In order to determine tissue specific gene expression, reverse transcription-PCR (RT-PCR) analyses were conducted using the RNA extracted from foregut, midgut, hindgut, Malpighian tubules, fat body, tracheae and integument. Tissue dissection, extraction of RNA, and RT-PCR conditions were as described in Chapter 2. The sequences of the primers used in the RT-PCRs are shown in Table 4.1.

4.3 Results

LC-MS/MS analysis of PM preparations identified 82 proteins including 6 peritrophins, 2 chitin modifying enzymes, 54 putative digestive enzymes, 4 putative enzymes with nondigestive function, 13 other proteins with identified homologues in other insects and 3 proteins with no identifiable homologue in the NCBI database (Table 4.2).

Table 4.1 RT-PCR primers used in expression analysis of genes encoding peritrophic matrix proteins in *M. configurata*.

Primer	Sequence
McAST Fp	CTCATGGGCCAGCAGTCGTTC
McAST Rp	CTCTTCTTGGGGCAGGTCTTCA
McALP1 Fp	TGCACTGCCACCGCTTATTTGT
McALP1 Rp	GCGTGCGTGATGCGTGTCGTAGT
Mcβ1, 3GLU Fp	ACTTGGAGCTTGGGCGTGTCA
Mcβ1, 3GLU Rp	CAGGGCTTAGGGGCAGGGTTGTT
McCBD3P Fp	CTCCGAACCGAACTACGATGATGA
McCBD3P Rp	TGGTGGTAGGCGCTGATGTAAATG
McCDA1 Fp	CGAGGCTTTCTTGTCAGCTTTC
McCDA1 Rp	TTGTCCAAGAGGGTTGCCGAG
McCDA2 Fp	CGTGCTCCCTTCGGTTTCTAT
McCDA2 Rp	ACTCCATTCTCTGCAAGGTTTCTC
McCLECT Fp	TATTTTGTGTTGCTGTTCGGTTTA
McCLECT Rp	AGCGGTTCATCAAGGTTCTCA
McdsRNase Fp	CCAACCACGAAGCACAAGGAACTA
McdsRNase Rp	TTACGGCAACACCAGGGAAAAAG
McMG176 Fp	GCTGATTGACGAGGCTGAGGAG
McMG176 Rp	CGGGGTTGCGTTCTGTGAGG
McIIM4 Fp	CAACATCAGCACCTGACTGC
McIIM4 Rp	CTCAGAGCAGATGTTAAGACGA
McPAN Fp	TTCAGTGGCGTAAGGTCGTTCTC
McPAN Rp	GGCATTGTTAGTGTCCAGGTCAGC
McPM1 Fp	CGCTGTAGATAACTCTGAAGGTG
McPM1 Rp	GGGAGCTTCTTCAGGATTACAGTTAC
McPOL Fp	CTCGTAGCAACAAGACAGGAAATA
McPOL Rp	TAAAAGTAACACGAGTAAAAATG
McPPAD1 Fp	CCACCGTCCTGCTCCTGAAC
McPPAD1 Rp	CAGGTCTGCAGTTGGAAGTC
McREPAT Fp	TCCTCTTTTACGACCGACTG

Table 4.1 (continued)

Primer	Sequence
McREPAT Rp	TGATTACCATTTACAAACGACCAA
McTUB Fp	AGCGTACCATCCAGTTCGTG
McTUB Rp	TGGCGTACATGAGGTCGAAC
357-80-1 Fp	CCGTAGAAGAAGTCGTGGAGT
357-80-1 Rp	ACATGAGCGTAAGTCGTAAACA
357-94-1 Fp	CACTAATGCCTACGCTGCTC
357-94-1 Rp	ATATTCTTTCGCCTCACATC
363-100-1 Fp	ATGAGAACCTGAGCTGCGTATCG
363-100-1 Rp	AATTGGTGCTGAGGGGCGTTAG
530-224-1 Fp	ATATTGTTGAAAGTAAGGCTGTCT
530-224-1 Rp	CCTGAGTGGGTTTTAGTTATGTTA
530-247-1 Fp	TTTTCAGAGTTAGGCATCAATA
530-247-1 Rp	GTTCTCTACTCACGGGTTCG

Abbreviations: ALP: Alkaline phosphatase; AST, astacin; β1, 3GLU, β-1, 3-glucanase; CBD3P, chitin binding domain 3 protein; CDA, chitin deacetylase; CLECT, C type lectin; Fp, forward primer; MG176, Midgut 176; IIM, insect intestinal mucin; Mc, *M. configurata*; PAN, pantetheinase; POL, polycalin, PPAD, protein with peritrophin-A domain; REPAT, response to pathogen; Rp, reverse primer; TUB, tubulin.

4.3.1 One and Two Dimensional Separation of PM Proteins

1D PAGE analysis of the strongly associated PM protein fraction revealed 12 to 14 protein bands ranging from 24 to >200 kDa (Figure 4.1). The exclusion of NaCl from the extraction buffer resulted in an additional protein band migrating at approximately 55 kDa. In contrast, approximately 25 proteins were released by rinsing in Ringer's solution. This loosely associated fraction contained a greater proportion of lower molecular weight proteins.

2D PAGE analysis of strongly and loosely associated PM protein fractions revealed approximately 25 predominant protein spots (Figure 4.2).

Table 4.2 Proteins identified from the *M. configurata* peritrophic matrix by LC-MS/MS.

No/Name	Type	# of Pep.	Mow. Scorecov.	Seq.	Accession #	Domain (Pssm ID)
A. STRUCTURAL PROTEINS						
1. McPM1	**Nonmucin peritrophin**	**3**	**1929**	**10**	**AY277403**	**Peritrophin-A (**155025**)**
2. McPPAD1	Nonmucin peritrophin	1	47	10	GU596430	Peritrophin-A (155025)
3. McCBD3P	**Nonmucin peritrophin**	**1**	**61**	**4**	**HM357863**	**Chitin Binding Domain 3 (145950)**
4. McIIM1	Mucin peritrophin	13	3214	17	AY057052	Peritrophin-A (155025)+Mucin
5. McIIM2	**Mucin peritrophin**	**4**	**605**	**17**	**FJ670567**	**Peritrophin-A n (155025)+Mucin**
6. McIIM4	Mucin peritrophin	5	738	15	FJ670569	Peritrophin-A (155025)+Mucin
B. ENZYMES						
Chitin modifying enzymes						
7. McCDA1	**Chitin deacetylase**	**13**	**2224**	**38**	**EU660852**	**Polysaccharide deacetylase 1 (164326)**
8. McCDA2	Chitin deacetylase	12	4496	41	HM357864	Polysaccharide deacetylase 1 (164326)+Mucin
Digestive enzymes						
9. McSP1	**Serine protease**	**3**	**500**	**13**	**FJ205402**	**Trypsin (29152)**
10. McSP2	Serine protease	1	41	6	FJ205440	Trypsin (29152)
11. McSP3	**Serine protease**	**1**	**38**	**2**	**FJ205433**	**Trypsin (29152)**
12. McSP13	Serine protease	2	204	12	FJ205420	Elastase (29152)
13. McSP17	**Serine protease**	**4**	**423**	**12**	**FJ205400**	**Elastase (29152)**
14. McSP22	Serine protease	2	104	12	FJ205443	Elastase (29152)
15. McSP23	**Serine protease**	**10**	**890**	**31**	**FJ205424**	**Chymotrypsin (29152)**
16. McSP24	Serine protease	11	3752	30	FJ205398	Chymotrypsin (29152)
17. McSP25	**Serine protease**	**4**	**329**	**20**	**FJ205416**	**Chymotrypsin (29152)**
18. McSP27	Serine protease	13	2631	32	FJ205413	Chymotrypsin (29152)
19. McSP28	**Serine protease**	**2**	**888**	**13**	**FJ205404**	**Chymotrypsin (29152)**
20. McSP29	Serine protease	6	300	31	FJ205412	Chymotrypsin (29152)
21. McSP30	**Serine protease**	**8**	**606**	**20**	**FJ205436**	**Chymotrypsin (29152)**
22. McSP32	Serine protease	5	209	17	FJ205441	Chymotrypsin (29152)
23. McSP33	**Serine protease**	**3**	**82**	**8**	**FJ205415**	**Elastase (29152)**
24. McSP34	Serine protease	2	125	11	FJ205445	Trypsin (29152)
25. McSP35	**Serine protease**	**5**	**505**	**18**	**FJ205405**	**Elastase (29152)**
26. McSP38	Serine protease	6	5120	39	FJ205426	Elastase (29152)
27. McSP48	**Serine protease**	**5**	**336**	**28**	**FJ205437**	**Chymotrypsin (29152)**
28. McSP51	Serine protease	8	1223	43	HM990173	Chymotrypsin (29152)
29. McSP52	**Serine protease**	**5**	**322**	**17**	**HM990174**	**Trypsin (29152)**
30. McSP53	Serine protease	4	250	14	HM990175	Chymotrypsin (29152)
31. McSP54	**Serine protease**	**3**	**203**	**16**	**HM990176**	**Chymotrypsin (29152)**
32. McSP55	Serine protease	8	759	34	HM990177	Chymotrypsin (29152)
33. McSP56	**Serine protease**	**5**	**372**	**18**	**HM990178**	**Chymotrypsin (29152)**
34. McSP57	Serine protease	4	379	22	HM990179	Trypsin (29152)
35. McSP58	**Serine protease**	**3**	**263**	**13**	**HM990180**	**Chymotrypsin (29152)**
36. McSP59	Serine protease	3	141	11	HM990181	Elastase (29152)
37. McSP60	**Serine protease**	**1**	**60**	**2**	**HM990182**	**Trypsin (29152)**
38. McSP61	Serine protease	9	1145	37	HM990183	Chymotrypsin (29152)
39. McSP62	**Serine protease**	**1**	**69**	**5**	**HM990184**	**Trypsin (29152)**
40. McSP64	Serine protease	3	258	15	HM990186	Elastase (29152)
41. McSP65	**Serine protease**	**4**	**390**	**13**	**HM990187**	**Elastase (29152)**
42. McAPN1	Aminopeptidase	5	263	11	HM357837	Peptidase M1 (144871)
43. McAPN3	**Aminopeptidase**	**24**	**3209**	**36**	**HM357833**	**Peptidase M1 (144871)+PepN (30656)**
44. McAPN4	Aminopeptidase	10	725	20	HM357829	Peptidase M1 (144871)
45. McAPN6	**Aminopeptidase**	**15**	**2665**	**35**	**HM357830**	**Peptidase M1 (144871)+ PepN (30656)**
46. McAPN7	Aminopeptidase	34	5542	37	HM357836	Peptidase M1 (144871)
47. McAPN8	**Aminopeptidase**	**2**	**202**	**5**	**HM357839**	**Peptidase M1 (144871)**
48. McCPB1	Carboxypeptidase	4	163	15	FJ210818	Peptidase M14 Carboxypeptidase (133072)
49. McCPA2	**Carboxypeptidase**	**6**	**325**	**9**	**FJ210819**	**Peptidase M14 Carboxypeptidase (175235)**
50. McCPB4	Carboxypeptidase	7	205	21	FJ210817	Peptidase M14 Carboxypeptidase (175235)
51. McCPA5	**Carboxypeptidase**	**2**	**134**	**11**	**HM357840**	**Peptidase M14 Carboxypeptidase (175235)**
52. McCPA6	Carboxypeptidase	1	61	5	HM357841	Peptidase M14 Carboxypeptidase (175235)
53. McIIL1	**Insect intestinal lipase**	**1**	**42**	**7**	**EU660853**	**Esterase-lipase (175390)**
54. McIIL3	Insect intestinal lipase	7	663	31	EU660855	Esterase-lipase (175390)
55. McIIL4	**Insect intestinal lipase**	**6**	**5273**	**26**	**HM357823**	**Esterase-lipase (175390)**
56. McIIL5	Insect intestinal lipase	2	169	9	HM357824	Esterase-lipase (175390)
57. McIIL6	**Insect intestinal lipase**	**1**	**79**	**2**	**HM357825**	**Esterase-lipase (175390)**
58. McIIL7	Insect intestinal lipase	2	177	6	HM357826	Esterase-lipase (175390)
59. McIIL8	**Insect intestinal lipase**	**1**	**53**	**8**	**HM357827**	**Esterase-lipase (175390)**
60. McIIL9	Insect intestinal lipase	1	52	6	HM357828	Esterase-lipase (175390)
61. Mcβ1, 3GLU	**Beta-1,3-glucanase**	**1**	**55**	**4**	**HM357842**	**Glyco hydrolase 16 (153598)**
62. McAMY	Alpha-amylase	1	34	1	HM357843	α-amylase (164103)+ α-amylase C (155067)
Other Enzymes						
63. McALP1	**Alkaline phosphate**	**5**	**172**	**14**	**HM357865**	**AlkPPc (175394)**

Table 4.2 (continued)

No/Name	Type	# of Pep.	Mow. Scorecov.	Seq.	Accession #	Domain (Pssm ID)
64. McdsRNase	dsRNase	11	838	22	HM357845	DNA/RNA nonspecific endonuclease (28975)
65. McAST	**Astacin**	**2**	**62**	**7**	**HM357847**	**ZnMc astacin like (58580)**
66. McPAN	Pantetheinase	2	76	8	HM357848 -	
C. OTHER PROTEINS						
67. McMG176A	**HMG176 like**	**5**	**613**	**32**	**HM357849**	**-**
68. McMG176B	HMG176 like	4	512	32	HM357850	-
69. McMG176C	**HMG176 like**	**3**	**406**	**23**	**HM357851**	**-**
70. McMG176D	HMG176 like	2	115	10	HM357852	-
71. McMG176E	**HMG176 like**	**3**	**371**	**27**	**HM357853**	**-**
72. McMG176F	HMG176 like	3	60	27	HM357854	-
73. McPOL	**Polycalin**	**20**	**4140**	**23**	**HM357855**	**Lipocalin (174555)**
74. McREPAT2	Repat	2	86	9	HM357858	-
75. McSerpin1A	**Serpin**	**2**	**42**	**24**	**AY148483**	**SERPIN (174005)**
76. McCLECT	C-type lectin	1	58	8	HM357859	C-type lectin (153057)
77. 363-100-1	**Lsti99 and Lsti201-like**	**3**	**322**	**15**	**HM357856**	**-**
78. 530-224-1	Lsti99 and Lsti201-like	1	779	3	HM357857	-
79. GmLe1	**Soybean lectin**	**3**	**73**	**7**	**K00821.1**	**L-type lectin (173887)**
D. PROTEINS WITH NO KNOWN HOMOLOGUES						
80. 357-94-1	-	**8**	**5959**	**35**	**HM357860**	**-**
81. 530-247-1	-	4	576	13	HM357861	-
82. 357-80-1	-	**1**	**50**	**4**	**HM357862**	**-**

Abbreviations: ALP: Alkaline phosphatase; AMY, α-amylase; APN, aminopeptidase-N; AST, astacin; β1, 3GLU, β-1, 3-glucanase; CBD3P, chitin binding domain 3 protein; CDA, chitin deacetylase; CLECT, C type lectin; CPA, carboxypeptidase A; CPB, carboxypeptidase B; dsRNase, double-strand RNase; GmLe1, soybean lectin; HMG176, *Helicoverpa* midgut 176; IIL, insect intestinal lipase; IIM, insect intestinal mucin; Mc, *M. configurata*; MG176, midgut 176; Mow., mowse; PAN, pantetheinase; Pep, peptide; PM, peritrophic matrix; POL, polycalin, PPAD, protein with peritrophin-A domain; Pssm, position specific scoring matrix; REPAT, response to pathogen; Seq. cov., Sequence coverage; SP, serine protease; TUB, tubulin; #, number.

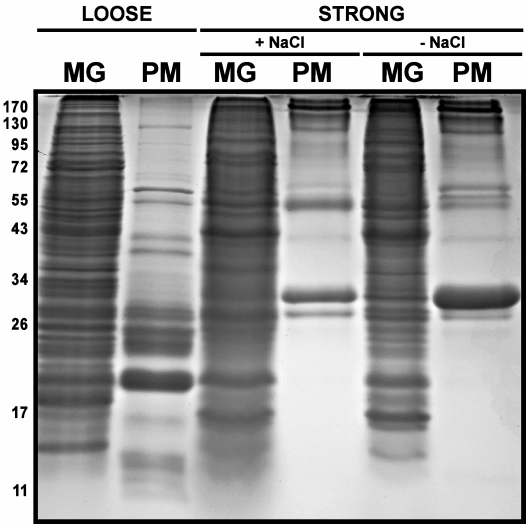


Figure 4.1 One dimensional denaturing gel electrophoresis of proteins loosely (LOOSE) and strongly (STRONG) associated with the peritrophic matrix (PM) and midgut (MG) of 5th instar *Mamestra configurata* larvae fed on artificial diet. Loosely associated proteins solubilized with Ringer's physiological solution and strongly associated proteins solubilized with 2.5% SDS, 5% β-mercaptoethanol with or without 500 mM NaCl at 100°C for 5 min are shown. Molecular weight markers (kDa) are shown in the left hand margin.

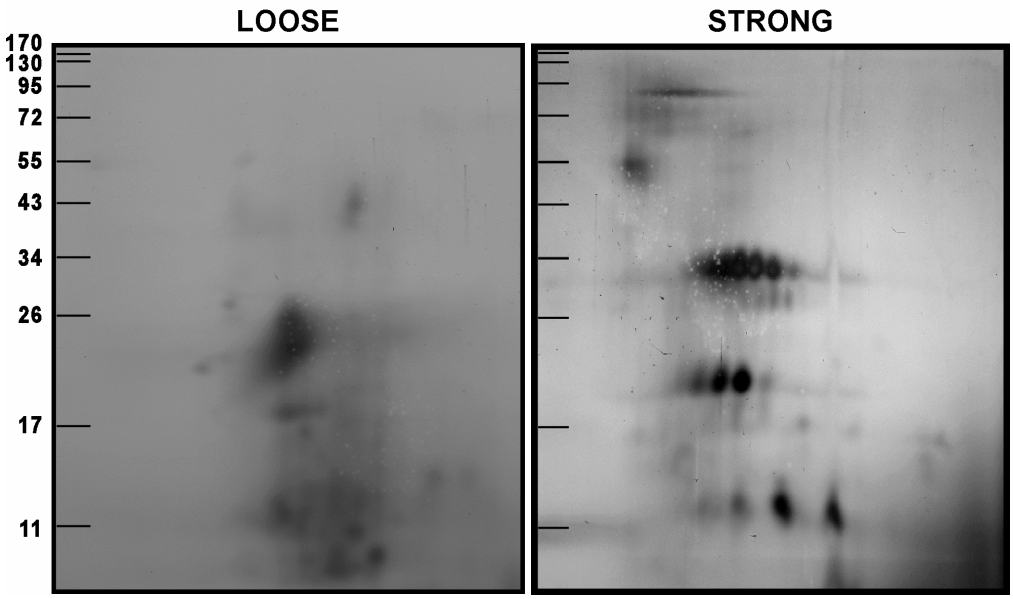


Figure 4.2 Two dimensional gel electrophoresis of loosely (LOOSE) and strongly (STRONG) associated peritrophic matrix proteins from 5th instar *Mamestra configurata* larvae. Loosely associated proteins solubilized by Ringer's physiological solution and strongly associated proteins solubilized with 2.5% SDS, 5% β-mercaptoethanol and 500 mM NaCl at 100°C for 5 min. Molecular weight markers (kDa) are shown in the left hand margin.

4.3.2 Peritrophins

Proteomic analysis of the PM revealed three nonmucin proteins (McPM1, McPPAD1 and McCBD3P) and three IIMs (McIIM1, 2, and 4). Two of these, McPM1 and McIIM1 had been previously identified (Shi et al., 2004).

Mascot analysis identified a single peptide hit linked to a putative 221 amino acid protein encoded by a 951 bp cDNA (Table 4.2, Appendix A). A conserved domain search (Marchler-Bauer et al., 2009) revealed a PAD; therefore, the protein was denoted *M. configurata* protein 1 with peritrophin-A domain (McPPAD1). McPPAD1 had a predicted molecular weight of 22.9 kDa after removal of the 18 amino acid signal peptide. Mascot analysis identified a single peptide hit linked to a putative 236 amino acid protein encoded by a 753 bp cDNA (Table 4.2, Appendix A). This was denoted *M. configurata* chitin binding domain 3 protein (McCBD3P) because a conserved domain search (Marchler-Bauer et al., 2009) revealed a Type 3 carbohydrate binding domain. CBD3 is composed of 110-117 amino acids containing a register of 10 cysteines with the consensus $C^1X_{12-13}C^2X_2C^3X_{51-52}C^4X_{9-12}C^5X_{35-36}C^6X_2C^7X_{14}C^8X_{8-9}C^9X_8C^{10}$. McCBD3P had a predicted molecular weight of 24.8 kDa without its 18 amino acid

signal peptide. Four peptide hits were identified by Mascot analysis that were linked to a putative 309 amino acid protein encoded by a 1077 bp cDNA (Table 4.2, Appendix A). This was denoted *M. configurata* IIM2 (McIIM2) because numerous amino acids were predicted to be *O*-glycosylated, an indication of a MD in addition to the two PADs that were identified by a conserved domain search (Marchler-Bauer et al., 2009). McIIM2 had a predicted molecular weight of 31.5 kDa without the 18 amino acid signal peptide. Five peptide hits linked to a putative 428 amino acid protein encoded by a 1335 bp cDNA (Table 4.2, Appendix A) were identified by Mascot analysis. The protein was predicted to have numerous *O*-glycosylated amino acids in $T^{198-214}$ and $T^{304-428}$ in addition to the four PADs identified by conserved domain search (Marchler-Bauer et al., 2009), and therefore was denoted *M. configurata* IIM4 (McIIM4). McIIM4 constituted only a partial sequence and a signal peptide was not predicted.

McPM1 and *McPPAD1* were expressed in foregut, hindgut, Malpighian tubules, tracheae, integument and fat body whereas *McCBD3P* was expressed in tracheae and very slightly in Malpighian tubules in addition to midgut (Figure 4.3). *McIIM4*, a typical IIM gene, was expressed only in midgut.

4.3.3 Enzymes

4.3.3.1 Chitin deacetylases

Proteomic analysis revealed that two CDAs were associated with the *M. configurata* PM. Mascot analysis identified 12 peptide hits linked to a putative 390 amino acid protein encoded by a 1258 bp cDNA and 13 peptide hits linked to a putative 425 amino acid protein encoded by a 1387 bp cDNA (Table 4.2, Appendix A). A conserved domain search (Marchler-Bauer et al., 2009) revealed a putative polysaccharide deacetylase domain for each protein; therefore, the proteins were denoted *M. configurata* chitin deacetylase 1 (McCDA1-390 amino acid protein) and chitin deacetylase 2 (McCDA2-425 amino acid protein). McCDA1 and McCDA2 had predicted molecular weights of 42.2 and 45.2 kDa, respectively, after removal of the 18 and 21 amino acid signal peptides. McCDA2 had 10 putative *O*-glycosylated amino acids in T^{27-54} together with 5 prolines. *McCDA1* was expressed primarily in midgut and weakly in foregut and tracheae. In contrast, *McCDA2* expression was specific to midgut (Figure 4.3).

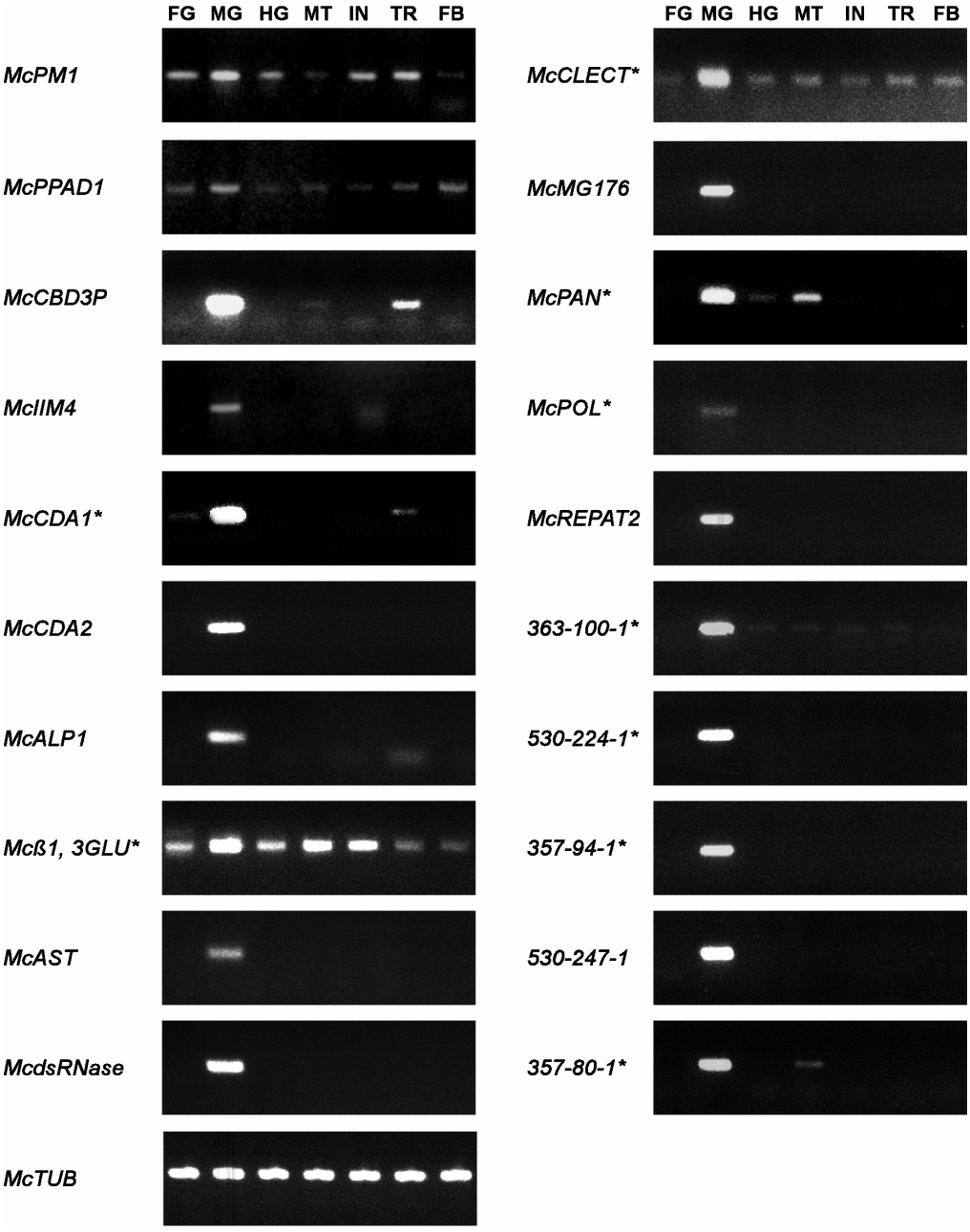


Figure 4.3 Expression of *Mamestra configurata* genes encoding peritrophic matrix proteins. Gene expression patterns were checked using total RNA from foregut (FG), midgut (MG), hindgut (HG), Malpighian tubules (MT), fat body (FB), tracheae (TR) and integument (IN) from 5th instar larvae by RT-PCR analysis with 35 or 40(*) cycles. Amplification of the *M. configurata* tubulin (*McTUB*) gene was used as a control to verify that equivalent amounts of RNA were used in the RT-PCRs. Abbreviations: ALP, alkaline phosphatase; AST, astacin; β1, 3GLU, β-1, 3-glucanase; CBD3P, chitin binding domain 3 protein; CDA, chitin deacetylase; CLECT, C type lectin; MG176, midgut 176; IIM, insect intestinal mucin; Mc, *M. configurata*; PAN, pantetheinase; POL, polycalin, PPAD1, protein 1 with peritrophin-A domain; REPAT, response to pathogen.

4.3.3.2 Putative digestive enzymes

4.3.3.2.1 Serine proteases

Mascot analysis identified between 1-13 peptide hits each linked to 33 putative serine proteases (SPs) (Marchler-Bauer et al., 2009) (Table 4.2, Appendix A). These include: 8 trypsins (D^{189}) [McSP1, 2, 3, 34, 52, 57, 60, and 62], 16 chymotrypsins (S^{189}) [McSP23-25, 27-30, 32, 48, 51, 53-56, 58, and 61] and 9 elastases (G^{189}) [McSP13, 17, 22, 33, 35, 38, 59, 64 and 65] (Figure 4.4). Nineteen of these (McSPs denoted up to 50) have been previously identified (Hegedus et al., 2003; Erlandson et al., 2010). The remaining 14 SPs (McSPs denoted above 50) were identified in the current study. Each McSP had H^{57}, D^{102}, and S^{195} forming the catalytic triad (except McSP30, 32, 38, 54, and 57), a GXSXG motif (except McSP30, 32, 33, 38, 52, 54, 57, and 59) and five conserved cysteines, C^{58} (except McSP48) and $C^{154, 179, 191, 220}$ (except McSP38) (Figure 4.4).

4.3.3.2.2 Exopeptidases

Mascot analysis identified between 2 and 34 peptide hits linked to 6 putative aminopeptidases (with peptidase M1 and/or PepN domains) (Marchler-Bauer et al., 2009) (Table 4.2, Appendix A). Three of these (McAPN1, 3, 4) have been previously identified (Hegedus et al., 2003; Erlandson et al., 2010). The new aminopeptidases were denoted *M. configurata* aminopeptidase 6, 7 and 8 (McAPN6-8). Two of these, McAPN6 and McAPN8 had partial sequences with a 2814 bp cDNA encoding a 840 amino acid and a 1134 bp cDNA encoding a 373 amino acid protein, respectively. McAPN7 full length cDNA was 3388 bp encoding a 1036 amino acid protein with a predicted molecular weight of 112.9 kDa. All McAPNs (except McAPN4) were predicted to have 1-6 *N*-glycosylated asparagines, while McAPN4, 6, and 7 were predicted to have 4-37 *O*-glycosylated threonines or serines. A zinc binding motif (HEXXH) was present in all full length McAPNs. Glycosylphosphatidylinositol (GPI) modification sites were predicted for McAPN3, 4, 6, and 7 at A^{463}, S^{276}, D^{702}or and D^{1015}, respectively. These features are shown in an alignment between *M. configurata* and *Helicoverpa armigera* APNs (Figure 4.5).

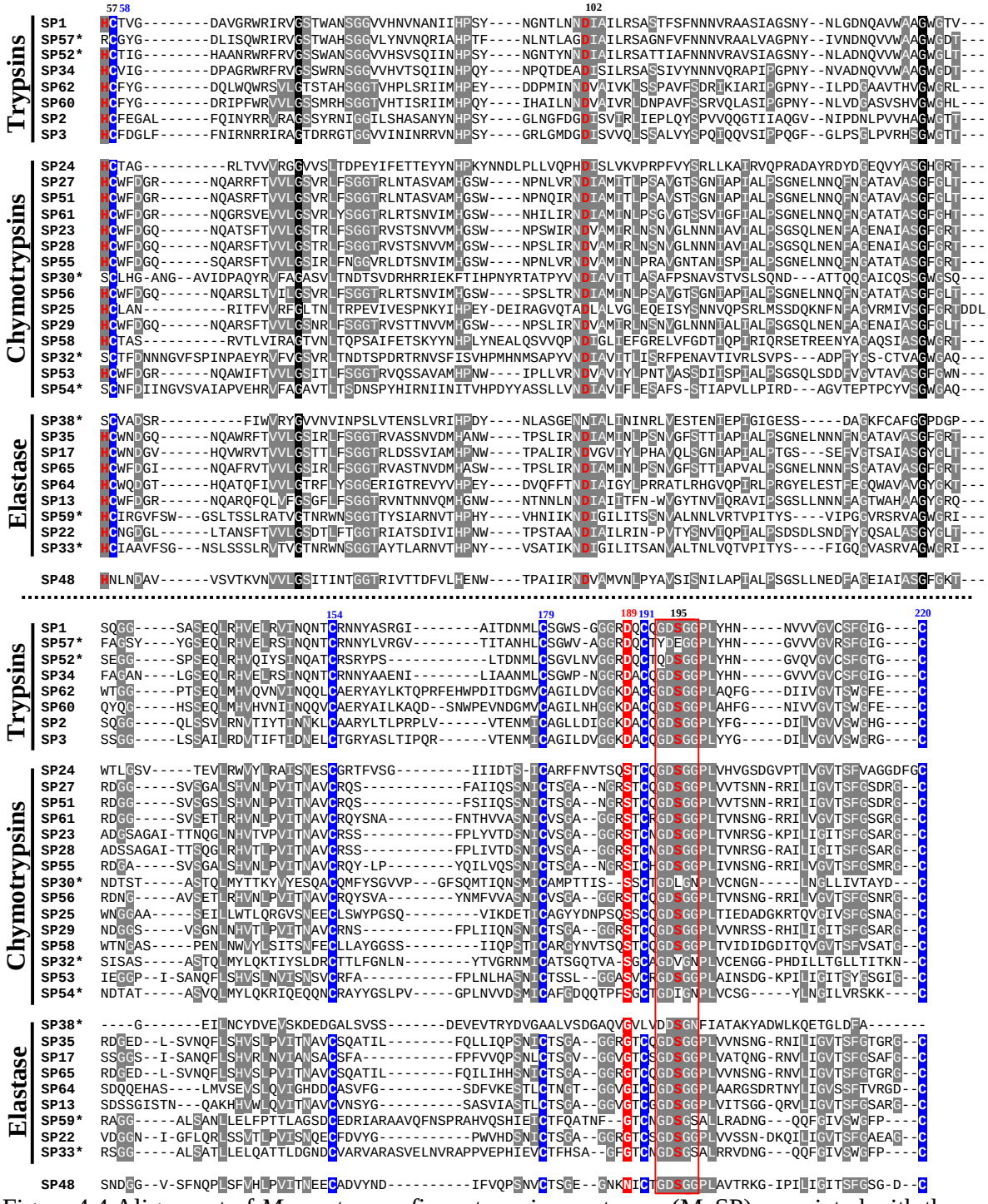
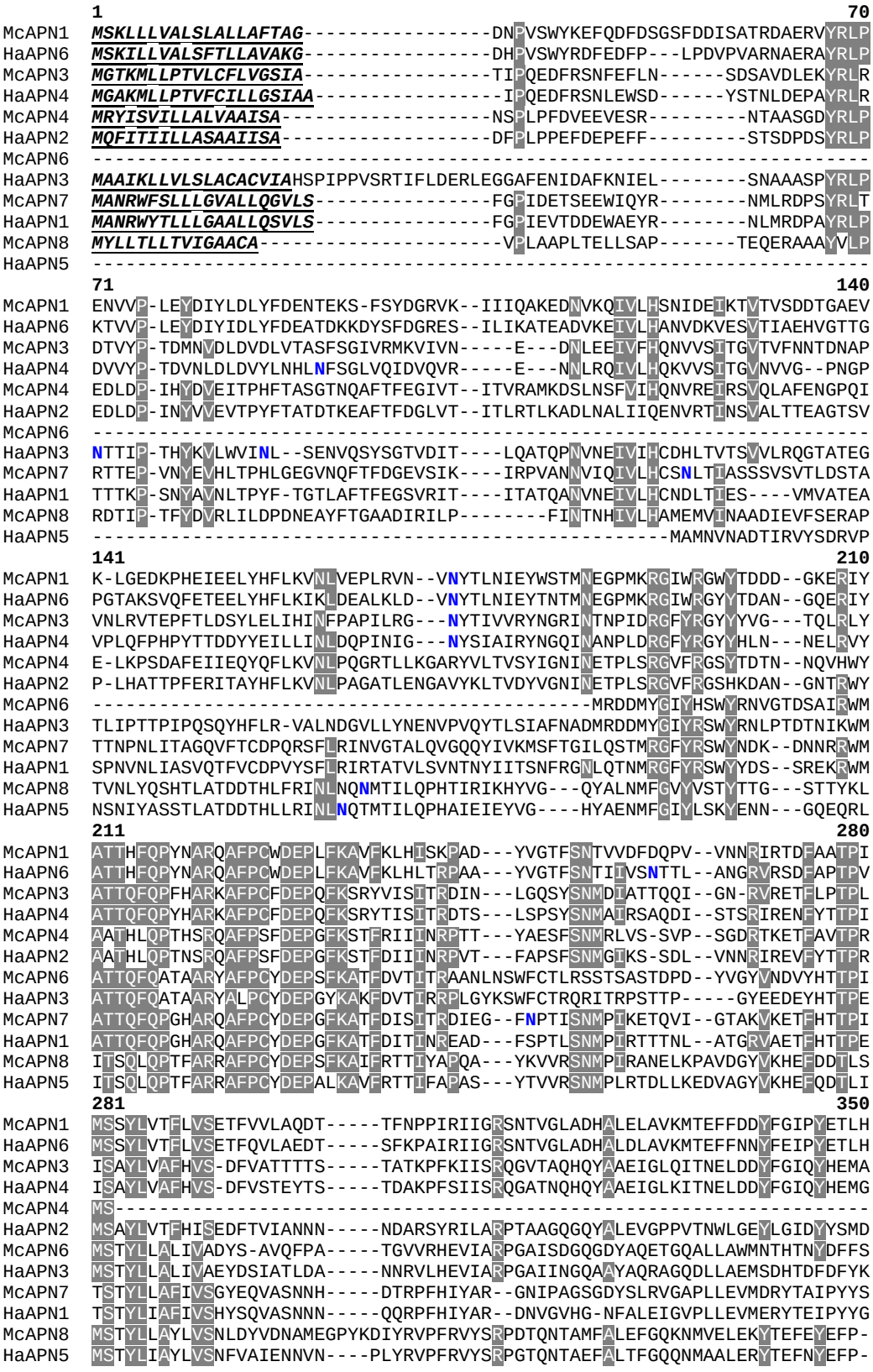
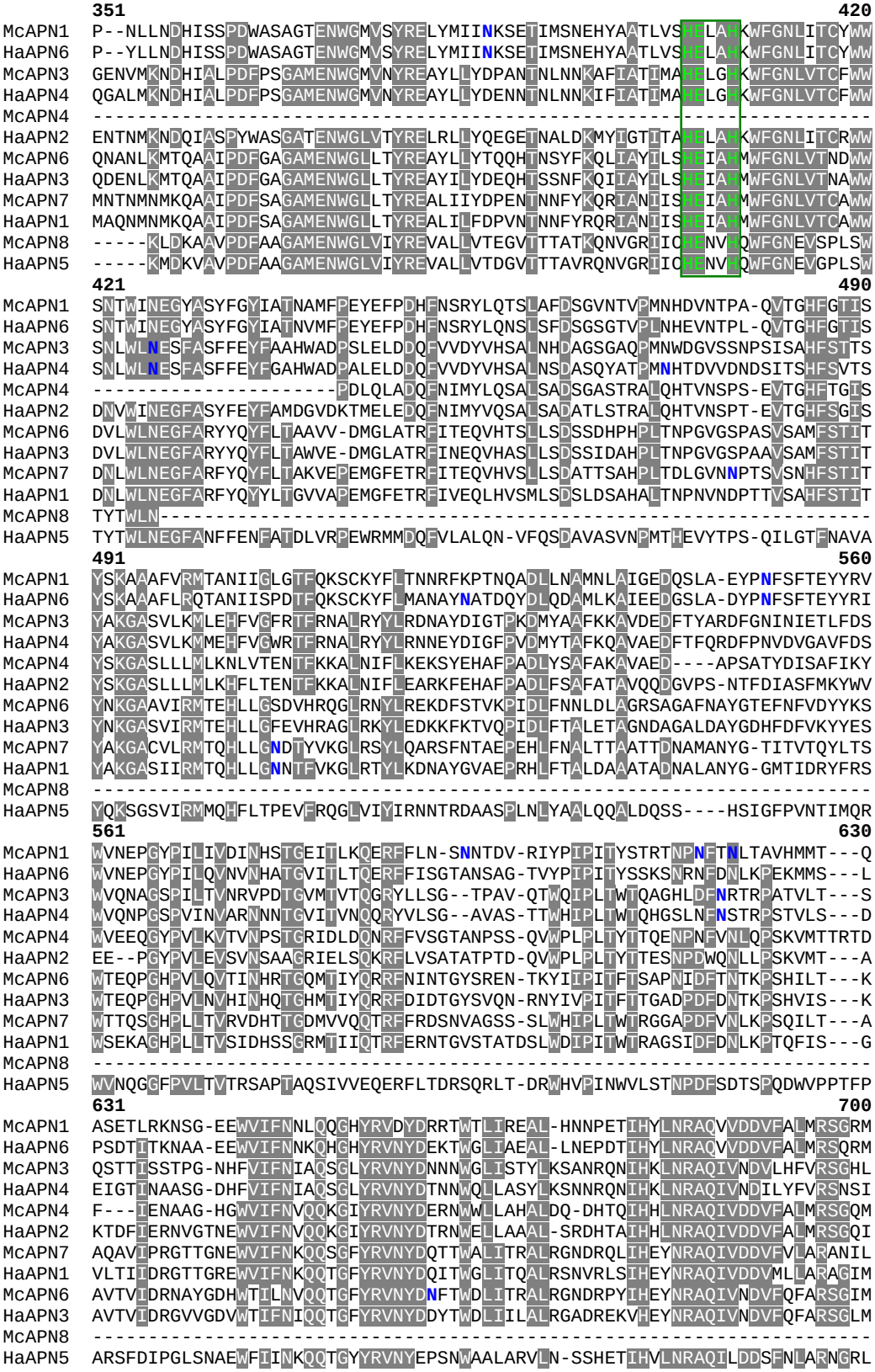


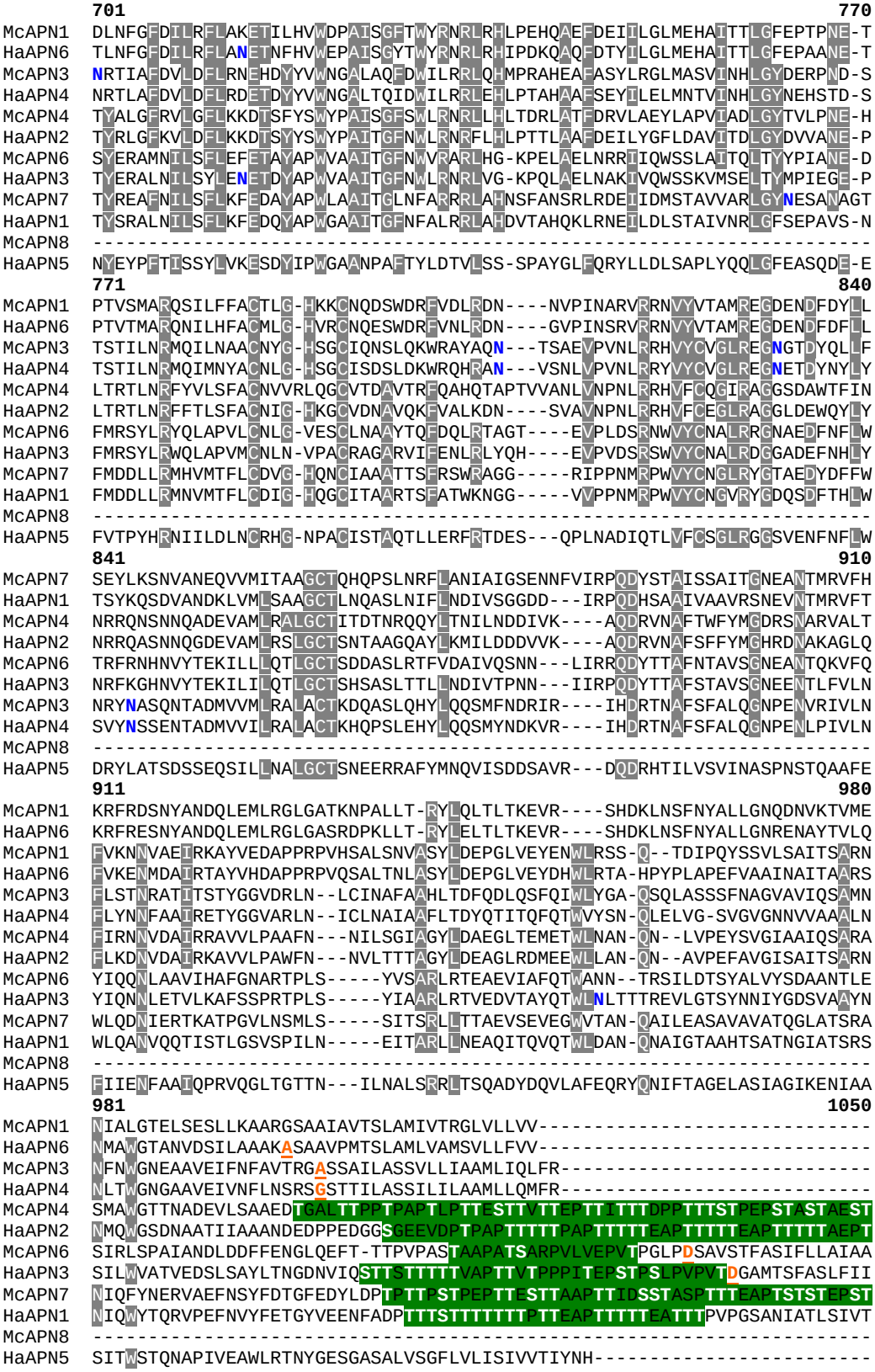
Figure 4.4 Alignment of *Mamestra configurata* serine proteases (McSP) associated with the peritrophic matrix. Locations of the amino acids present in all (black background) or the majority (gray background), H^{57}, D^{102}, and S^{195} forming the catalytic triad (red lettering), GXSXG motif (red box), specificity amino acid at position 189 (red background) and conserved cysteines (blue background) are shown. Those McSPs denoted by "⋆"may be inactive due to lack of proper GXSXG motif.



(Figure 4.5 continued)



(Figure 4.5 continued)



(Figure 4.5 continued)

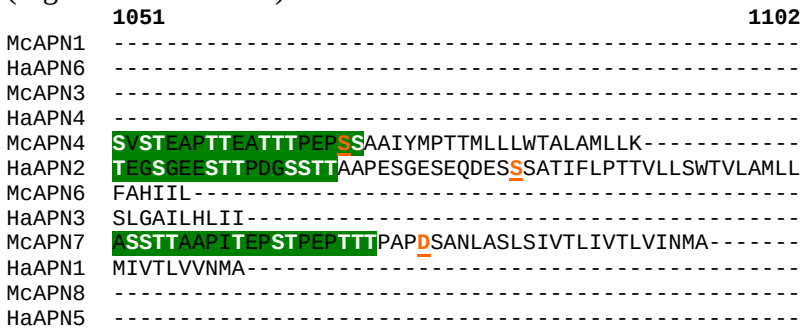


Figure 4.5 Alignment of *Mamestra configurata* peritrophic matrix aminopeptidases (APNs) with those from *H. armigera*. Location of the signal peptides (bold, italics and underlined), zinc binding site "HEXXH" (green lettering in green box), putative glycosyl phosphatidylinositol modification sites (orange lettering, underlined), asparagines that were predicted to be *N*-glycosylated (blue lettering), or *O*-glycosylated amino acids (white lettering) within the serine-threonine-proline rich region (green background) and amino acids present in the majority (gray background are shown. Accession numbers of the HaAPNs are: HaAPN1 (ACC68683), HaAPN2 (AAP37951), HaAPN3 (AAP37953), HaAPN4 (AAP37950), HaAPN5 (ABN64103), and HaAPN6 (ACA35025).

Mascot analysis identified one to seven peptide hits linked to five putative carboxypeptidases (with a peptidase M14 carboxypeptidase domain) (Marchler-Bauer et al., 2009) (Table 4.2, Appendix A). These include three A Type carboxypeptidases, CPA, (F^{255}, V^{255} and S^{255}) [McCPA2, 5 and 6, respectively] and two B Type carboxypeptidases, CPB, (R^{255} and E^{255}) [McCPB1 and CPB4, respectively] (Figure 4.6). McCPB1, McCPA2 and McCPB4 were identified previously (Hegedus et al., 2003; Erlandson et al., 2010). The two new carboxypeptidases, McCPA5 and CPA6, had partial sequences with a 665 bp cDNA encoding a 200 amino acid protein and a 682 bp cDNA encoding a 226 amino acid protein, respectively. All *M.configurata* carboxypeptidases that were found to be PM associated had a signal peptide, an arginine as a putative trypsin cleavage site near their amino termini, a zinc binding motif (H^{69}, E^{72} and H^{196}, numbering is according to the *Bos taurus* carboxypeptidases) (except McCPA5) and one or two *N*-glycosylated asparagines (except McCPA2). In addition, a pair of conserved cysteines (C^{138} and C^{161}) was found in McCPB1, McCPB4 and McCPA6; however, McCPA2 and McCPA5 lacked at least one of these. These features are shown in an alignment of McCPs with the model CPA (P00730) and CPB (P00732) from *B. taurus* (Figure 4.6).

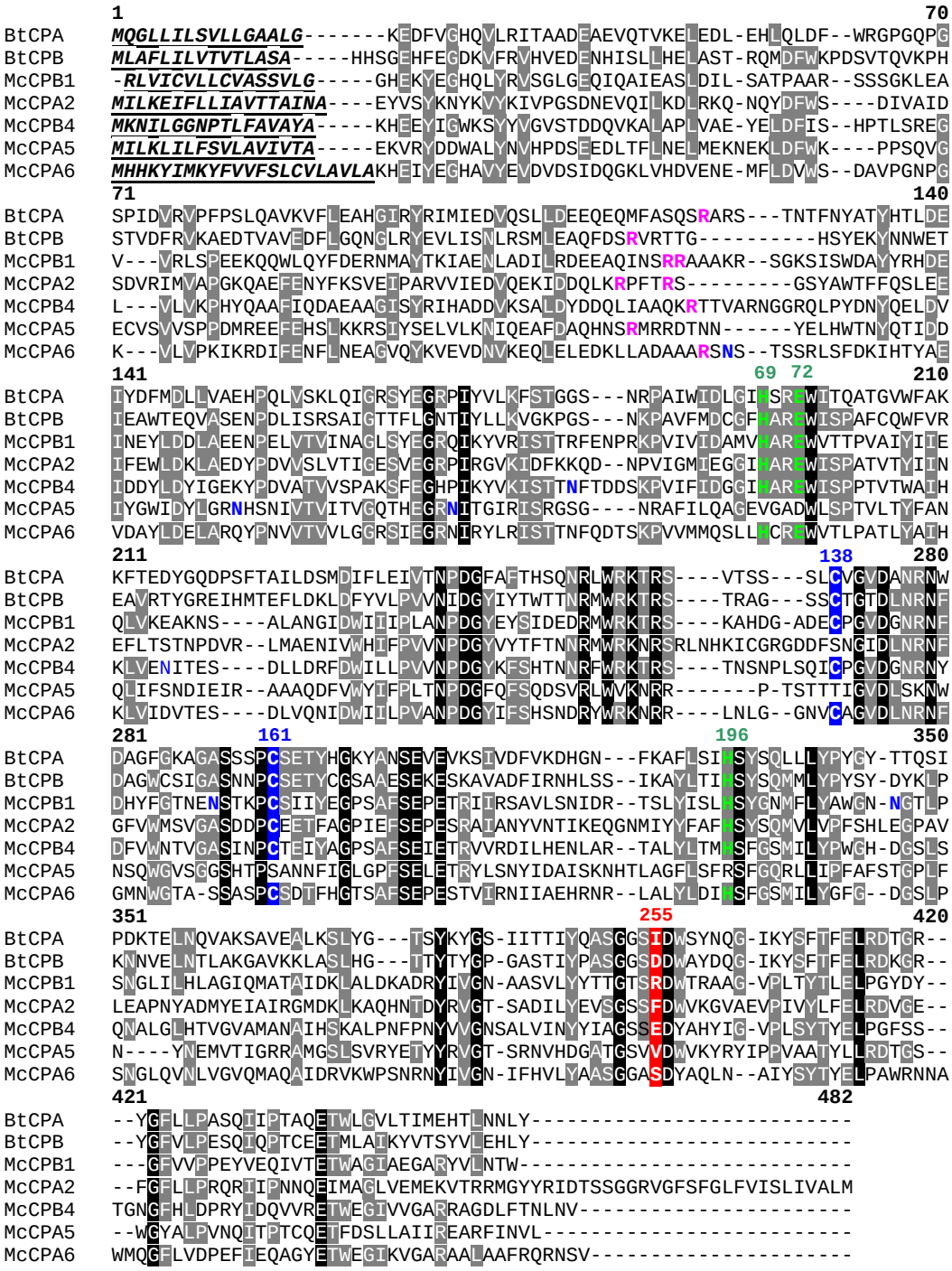


Figure 4.6 Alignment of *Mamestra configurata* peritrophic matrix associated carboxypeptidases with *Bos taurus* BtCPA and BtCPB. Locations of the signal peptides (bold, italics and underlined), putative trypsin cleavage sites (pink lettering), putative zinc binding amino acids (green lettering), cysteines that form putative disulphide bridges (blue background), specificity amino acid (red background), asparagines predicted to be *N*-glycosylated (blue lettering), amino acids present in all (black background) or the majority (gray background), are shown. GenBank accession numbers for BtCPA and BtCPB are P00730 and P00732, respectively.

4.3.3.2.3 Insect intestinal lipases

Mascot analysis identified one to seven peptide hits linked to eight putative proteins with an esterase-lipase domain (Marchler-Bauer et al., 2009) (Table 4.2, Appendix A). These were denoted *M. configurata* insect intestinal lipases 1, and 3-9 (McIIL1, McIIL3-9). McIIL1, 3, and 6 were predicted to have a signal peptide and a stop codon. The 1127 bp McIIL1 cDNA encoded a 331 amino acid protein with a predicted molecular weight of 33.4 kDa after removal of the 18 amino acid signal peptide. The 1042 bp McIIL3 cDNA encoded a 296 amino acid protein with a predicted molecular weight of 30.6 kDa after removal of the 16 amino acid signal peptide. The 1154 bp McIIL6 cDNA encoded a 332 amino acid protein with a predicted molecular weight of 33.7 kDa after removal of the 16 amino acid signal peptide. McIIL4, 5, 7, 8, and 9 were not predicted to have a signal peptide because they were only partial cDNA sequences ranging between 470 bp and 1080 bp in size. McIILs (except McIIL3 and McIIL4) contained S^{152}, D^{176}, and H^{263} forming the catalytic triad, and a GXSXG signature motif (Figure 4.7). All McIILs had a conserved arrangement of cysteines, $CX_{4-9}CX_{22-25}CX_{10}C$, near the carboxy terminus (Figure 4.7). Sequence information is not available for the McIIL8 or McIIL9 catalytic region or for the McIIL9 conserved cysteine rich region.

4.3.3.2.4 β-1, 3-Glucanase

Mascot analysis identified a single peptide hit linked to a 190 amino acid protein encoded by a 713 bp partially sequenced cDNA (Table 4.2, Appendix A). This was denoted *M. configurata* β-1, 3-glucanase (Mcβ1, 3GLU) because BLAST analysis revealed significant similarity (identity value>92%) with β1, 3GLUs from *Spodoptera littoralis* (EU770374), *H. armigera* (EF600056), and *S. frugiperda* (EF641300). Furthermore, a conserved domain search (Marchler-Bauer et al., 2009) identified a glycosyl hydrolase family 16 domain. Mcβ1, 3GLU was expressed in foregut, midgut, hindgut, Malpighian tubules, tracheae, integument and fat body (Figure 4.3).

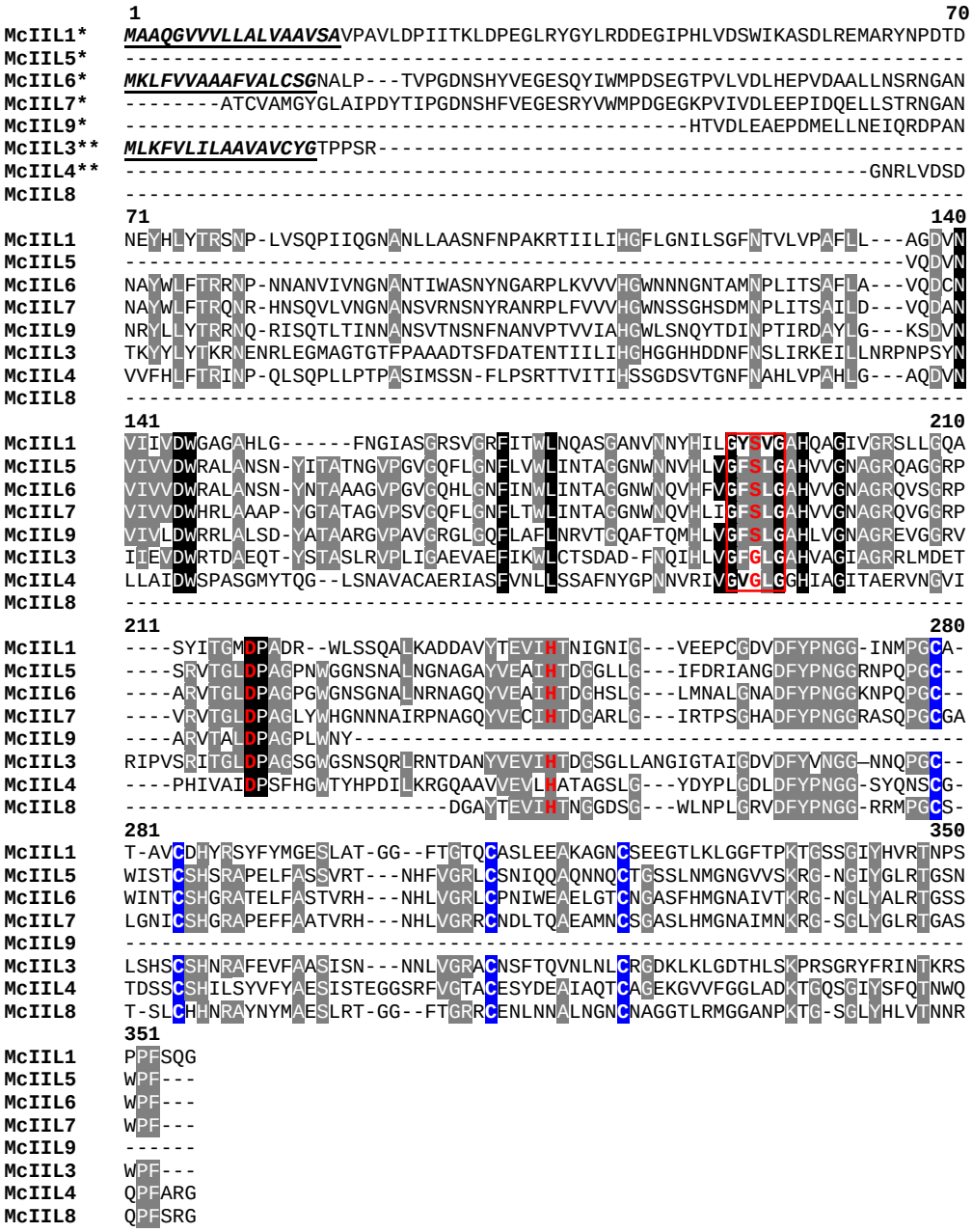


Figure 4.7 Alignment of *Mamestra configurata* insect intestinal lipases (McIIILs) associated with the peritrophic matrix. The locations of the signal peptides (bold, italics and underlined), S^{152}, D^{176}, and H^{263} forming the catalytic triad (red lettering), region containing the GXSXG motif (red box), conserved cysteines near the carboxy terminus (blue background), amino acids present in all (black background) or the majority (gray background) are shown. IILs that were predicted to be active (having at least one consensus catalytic site) are denoted by "*", whereas IILs that were predicted to be inactive (having a mutated catalytic site) are denoted by "**". The sequences have been deposited in GenBank as: McIIL1 (EU660853), McIIL3 (EU660855), McIIL4 (HM357823), McIIL5 (HM357824), McIIL6 (HM357825), McIIL7 (HM357826), McIIL8 (HM357827) and McIIL9 (HM357828).

4.3.3.2.5 α-Amylase

Mascot analysis identified a single peptide hit linked to a putative 500 amino acid protein encoded by a 1631 bp partially sequenced cDNA (Table 4.2, Appendix A). A conserved domain search (Marchler-Bauer et al., 2009) revealed an *α*-amylase domain; therefore, the protein was denoted *M. configurata* α-amylase (McAMY). McAMY had a predicted molecular weight of 54.5 kDa after removal of the 16 amino acid signal peptide and contained 10 cysteines ($C^{44, 102, 155, 172, 380, 386, 422, 445, 452, 464}$).

4.3.3.3 Other Enzymes

4.3.3.3.1 Alkaline phosphatase

Mascot analysis produced five peptide hits linked to a 469 amino acid protein encoded by a 1718 bp partial cDNA sequence (Table 4.2, Appendix A). A conserved domain search (Marchler-Bauer et al., 2009) revealed an alkaline phosphatase (alkPPc) domain; therefore, the protein was denoted *M. configurata* alkaline phosphatase 1 (McALP1). McALP1 was predicted to contain a GPI anchor at S^{445} and three *N*-glycosylated asparagines at $N^{130, 225, 387}$. *McALP1* expression was restricted to midgut (Figure 4.3).

4.3.3.3.2 dsRNASE

Mascot analysis identified 11 peptide hits linked to a 444 amino acid protein encoded by a 1427 bp cDNA (Table 4.2, Appendix A). A conserved domain search (Marchler-Bauer et al., 2009) revealed a DNA/RNA nonspecific endonuclease domain and BLAST analysis identified *Bombyx mori* dsRNase as the top hit; therefore, the protein was denoted *M. configurata* dsRNase (McdsRNase). The predicted molecular weight of McdsRNase was 47.9 kDa after removal of the 16 amino acid signal peptide. *McdsRNase* expression was specific to midgut (Figure 4.3).

4.3.3.3.3 Astacin

Mascot analysis identified two peptide hits linked to a 371 amino acid protein encoded by a 1310 bp cDNA (Table 4.2, Appendix A). A conserved domain search (Marchler-Bauer et al., 2009) revealed a ZnMc astacin like domain. Therefore, the protein was denoted *M. configurata* astacin (McAST). McAST molecular weight was calculated as 41.6 kDa after removal of the 13 amino acid signal peptide. McAST had two R-S cleavage sites in the amino terminus, a typical

zinc binding motif "HEXXH" and four cysteines at $C^{124, 148, 174, 284}$. *McAST* expression was specific to midgut (Figure 4.3).

4.3.3.3.4 Pantetheinase

Mascot analysis produced two peptide hits linked to a 216 amino acid protein encoded by a 1060 bp partially sequenced cDNA (Table 4.2, Appendix A). BLAST analysis of the protein revealed various pantetheinases (vanin-like proteins) from *Nasonia vitripennis* (XM_001603761), *Tribolium castaneum* (EFA12337) and *Culex quinquefasciatus* (XM_001849712). The protein was denoted *M. configurata* pantetheinase (McPAN). McPAN was predicted to have 20 *O*-glycosylated amino acids in T^{151}-T^{188}. *McPAN* was expressed in hindgut and Malpighian tubules in addition to midgut (Figure 4.3).

4.3.4 Other proteins

4.3.4.1 McMG176

Mascot analysis identified two to five peptide hits linked to six related proteins encoded by cDNAs varying between 471-686 bp in length (Table 4.2, Appendix A). BLAST analysis indicated that all these cDNAs encoded proteins with identity to MG176 protein from *H. armigera* (DQ847154). Due to the high degree of similarity among these proteins, they were collectively denoted *M. configurata* MG176 orthologs A-F (McMG176A-F). Only one of the cDNAs (McMG176F) was full length (686 bp), encoding a 163 amino acid protein with a predicted molecular weight of 17.5 kDa after the removal of 20 amino acid signal peptide. McMG176 orthologs were predicted to have a transmembrane domain at the carboxy terminus (amino acids 128-150). *McMG176* expression was detected only in midgut (Figure 4.3).

4.3.4.2 Lipocalin

Mascot analysis identified 20 peptide hits linked to a 1111 amino acid protein encoded by a 3520 bp cDNA (Table 4.2, Appendix A). A conserved domain search revealed seven polymeric lipocalin domains (Marchler-Bauer et al., 2009). This protein was denoted *M. configurata* polycalin (McPOL). McPOL has a predicted molecular weight of 119 kDa after removal of the 20 amino acid signal peptide. *McPOL* expression was specific to midgut (Figure 4.3).

4.3.4.3 REPAT

Mascot analysis identified two peptide hits linked to a 124 amino acid protein encoded by a 499 bp cDNA (Table 4.2, Appendix A). BLAST analysis revealed that the protein was an ortholog of *Spodoptera exigua* response to pathogen protein 2, SeREPAT2 (ACI90727). The protein was denoted *M. configurata* REPAT2 (McREPAT2). The McREPAT2 had a predicted molecular weight of 11.3 kDa after the removal of the 19 amino acid signal peptide. *McREPAT2* was expressed only in midgut (Figure 4.3).

4.3.4.4 Serpin

Mascot analysis identified two peptide hits linked to a protein that was identified as *M. configurata* Serpin1A (McSerpin1A). Expression of *McSerpin1A* in the midgut and its localization to the PM was demonstrated previously (Chamankhah et al., 2003; Hegedus et al., 2008) (Table 4.2, Appendix A).

4.3.4.5 C type lectin

Mascot analysis identified one peptide hit linked to a 219 amino acid protein encoded by a 667 bp partial cDNA sequence. A conserved domain search (Marchler-Bauer et al., 2009) revealed two C type lectin domains, a type of carbohydrate recognition domain (CRD). The protein was denoted *M. configurata* C type lectin (McCLECT). McCLECT was predicted to have a 19 amino acid signal peptide, four Ca^{2+} binding sites, two *N*-glycosylated asparagines (N^{75} and N^{185}) and eight cysteines ($C^{23, 51, 125, 139, 147, 160, 176, 193}$) in which the first five cysteines were found in the first C type lectin domain. Expression of *McCLECT* was detected in foregut, hindgut, Malpighian tubules, tracheae, integument and fat body in addition to midgut (Figure 4.3).

4.3.4.6 Lsti99 and Lsti201-like proteins

Mascot analysis identified one to three peptide hits linked to two putative proteins encoded by a 748 bp (363-100-1) and a 1060 bp (530-224-1) partial cDNA sequence (Table 4.2, Appendix A). BLAST analysis indicated that both proteins shared identity with two PM proteins from *Loxostege sticticalis*, Lsti99 (FJ798745) and Lsti201 (FJ798746) (Figure 4.8). 363-100-1 (247 amino acids) was predicted to have a 20 amino acid signal peptide; however, the partial

cDNA lacked the region encoding the carboxy terminus. 530-224-1 (143 amino acids) sequence corresponded to an internal region when aligned with the Lsti99 or Lsti201. The *L. sticticalis* proteins have eight conserved cysteine registers of which seven are found in 363-100-1 (Figure 4.8). *363-100-1* was expressed primarily in midgut, and to a lower extent in hindgut, Malpighian tubules and tracheae whereas *530-224-1* was expressed only in midgut (Figure 4.3).

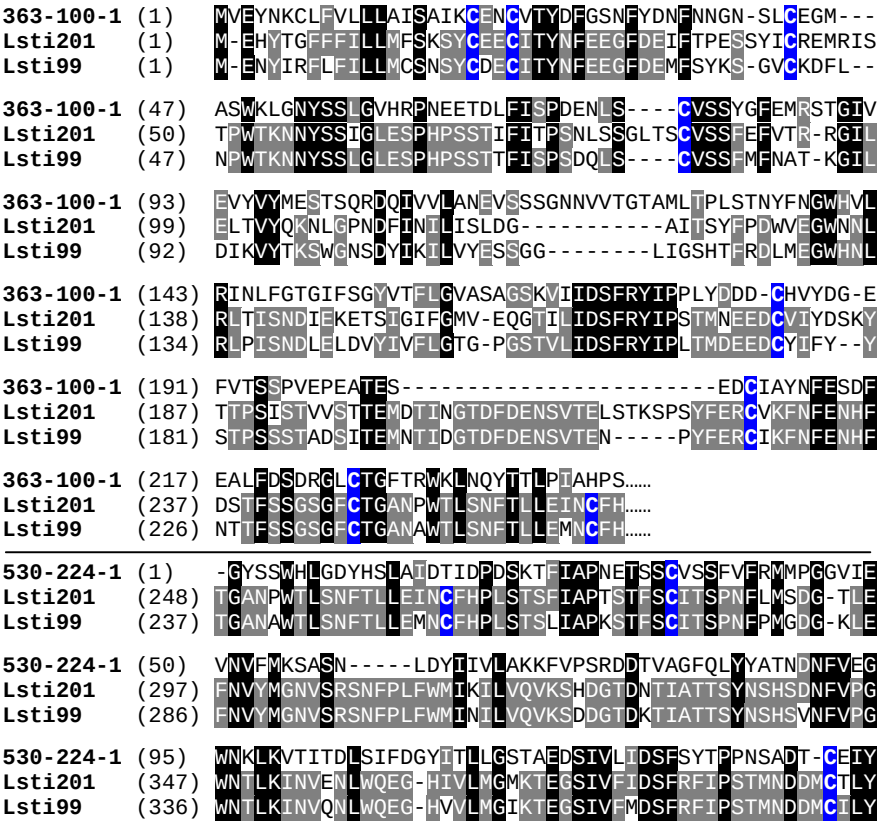


Figure 4.8 Alignment of *Mamestra configurata* peritrophic matrix proteins 363-100-1 and 530-224-1 with *L. sticticalis* Lsti99 and Lsti201. Amino acids present in all (black background) or the majority (gray background), and cysteines (blue background) are shown.

4.3.4.7 Soybean lectin (GmLe1)

Mascot analysis revealed three peptide hits linked to a protein that was previously identified as soybean lectin (GmLe1) (Shi et al., 2004) (Table 4.2, Appendix A).

4.3.5 Proteins without Orthologs

Mascot analysis identified one to eight peptide hits linked to three putative proteins, 357-94-1 (188 amino acids), 530-247-1 (209 amino acids) and 357-80-1 (143 amino acids). A search of the NCBI database using BLAST did not reveal any proteins related to these (Table 4.2,

Appendix A). All proteins were predicted to have a signal peptide. 357-94-1 and 530-247-1 were quite similar (identity value: 88%) and had eight conserved cysteines with an arrangement of $C^1X_{17}C^2X_9C^3X_1C^4X_{11}C^5X_{46\text{-}48}C^6X_{10}C^7X_{22}C^8$ (the last cysteine is not available due to missing sequence in 357-94-1). 357-80-1 was predicted to have seven *O*-glycosylated threonine or serines interspersed with prolines in $T^{88\text{-}112}$. Expression of *357-94-1* and *530-247-1* was restricted to midgut whereas expression of *357-80-1* was also detected in Malpighian tubules (Figure 4.3).

4.4 Discussion

Previous studies indicate that the protein component of the PM varies with respect to the number and type of proteins described among insect species (Adang and Spence, 1982). For example, two major proteins were reported from the PMs of adult *Simulium vitatum* (Ramos et al., 1994) and *Lutzomyia longipalpis* (Secundino et al., 2005). In contrast, the PM of adult *Glossina morsitans morsitans* had 40 proteins (Lehane et al., 1996), while *Aedes aegypti* and *Anopheles gambiae* had 20-40 major PM proteins (Moskalyk et al., 1996).

To date, only two studies have analyzed the PM protein content using an LC-MS/MS proteomics approach. Campbell et al. (2008) reported 43 proteins from the PM of larval *H. armigera* and Dinglasan et al. (2009) reported 209 proteins from the PM of *An. gambiae* adults, though the latter represented both loosely and strongly associated proteins. In my study, LC-MS/MS analysis of the *M. configurata* PM identified 82 proteins. The higher number of putative proteins associated with the *M. configurata* PM compared to that of *H. armigera* PM may reflect the different protein isolation protocols, or the longer liquid chromatography column runs and detection times that allowed me to develop a “deeper” proteomics dataset. As well, the *M. configurata* EST datasets used to search for peptide matches were more diverse than that used for *H. armigera* due to the inclusion of cDNA libraries from larvae fed on three diet regimens: artificial diet, artificial diet spiked with soybean trypsin inhibitor and canola.

4.4.1 Peritrophins

Peritrophins are essential structural components of the PM and have been identified in lepidopterans (Wang and Granados, 1997a), dipterans (Shen et al., 1999), coleopterans (Zhou et al., 2009) and hymenopterans (Marques-Silva et al., 2005). Proteomic analysis of the *M. configurata* larval PM revealed six peritrophins, of which McIIM1 and McPM1 had been

previously reported (Shi et al., 2004). All reported lepidopteran peritrophins have a PAD, which is a Type 2 CBD; however, McCBD3P has a Type 3 CBD, and this is the first report of such a domain in lepidopteran peritrophins. A Type 3 CBD was also recently reported from a dipteran peritrophin, AgAper25b, in *An. gambiae* (Dinglasan et al., 2009). Furthermore, several orthologs are present in other dipterans, suggesting this family of proteins may be a common component of the PM. As CBD3 contains 10 cysteines, as in the case of PCD, future classification should be based on structural information, not only the number of cysteines.

Peritrophins show variability in their expression patterns. *McPM1* and *McPPAD1* are expressed in foregut, hindgut, Malpighian tubules, tracheae, integument and fat body; whereas, *McCBD3P* is expressed in tracheae and Malpighian tubules in addition to the primary expression site, midgut (Figure 4.3). These findings clearly indicate that expression of peritrophin genes is not restricted to tissues synthesizing PM (i.e., the midgut). In contrast, *McIIM4*, a typical IIM, is expressed only in the midgut, which was also shown for *McIIM1* (Shi et al., 2004).

Many recombinant peritrophins have been demontrated to bind chitin *in vitro* (Elvin et al., 1996; Shen and Jacobs-Lorena, 1998; Wang et al., 2004a). In particular, nonmucin peritrophins with long chains of CBDs (e.g., McPM1 and TnCBP1) bind tightly to chitin and maintain a strong matrix via cross-linking and locking the chitin fibrils through CBD domains (Wang et al., 2004a). IIMs may also be involved in cross-linking of fibrils but are primarily proposed to lubricate the passage of food through the midgut, protect the PM from attack by midgut digestive proteases, and enable the midgut epithelium to avoid pathogen invasion (Wang and Granados, 1997a; b).

4.4.2 Enzymes

4.4.2.1 Chitin deacetylases

CDAs catalyze the release of acetyl groups from chitin to form chitosan (Dixit et al., 2008). Two CDAs, McCDA1 and McCDA2, are associated with the *M. configurata* PM. Interestingly, McCDA2 has a mucin like domain in the amino terminus. Two CDAs were also reported from *H. armigera* PM; however, neither of these had a mucin-like domain (Campbell et al., 2008). Midgut CDAs have also been reported from other lepidopterans (Guo et al., 2005; Nègre et al., 2006; Simpson et al., 2007), dipterans (Pauchet et al., 2009b) and coleopterans (Dixit et al., 2008). Expression of most of the genes encoding CDAs associated with PM was

reported to be restricted to midgut (Guo et al., 2005; Dixit et al., 2008; Arakane et al., 2009). Although *McCDA2* expression was midgut specific, *McCDA1* was also expressed weakly in foregut and tracheae. *HaCDA5A*, a gene encoding a PM associated CDA in *H. armigera*, was expressed in Malpighian tubules and fat body in addition to midgut (Jakubowska et al., 2010). The functions of insect CDAs, particularly the PM associated CDAs, is not well known, though roles in immunity and molting have been suggested (Dixit et al., 2008).

It is noteworthy that another group of chitin modifying enzyme, chitinases, are involved in PM degradation during molting and are associated with the PMs of *C. fumiferana* (Zheng et al., 2003) and *A. gambiae* (Dinglasan et al., 2009). A chitinase, McCHI, is also present in *M. configurata* midgut (Chapter 3); however, current proteomic analysis did not reveal its association with the PM. The PMs used in this analysis are from feeding larvae, and therefore, the abundance of *McCHI* may be too low.

Expression of chitinases in the midgut where the PM occurs, the presence of a chitin binding domain in the enzyme, and the fact that the PM contains chitin, suggest that they are physically associated with the PM. Such an association was demonstrated for Group I chitinases suggesting that members of this group may be involved in PM turn over.

4.4.2.2 Putative digestive enzymes

4.4.2.2.1 Serine proteases

Serine proteases are the major endopeptidases in the lepidopteran larval midgut (Christeller et al., 1992; Ferreira et al., 1994). A large suite (McSP1-50) of putative serine proteases have been identified from *M. configurata* larvae fed on three diets (Hegedus et al., 2003; Erlandson et al., 2010). The current PM proteome dataset includes 19 McSPs that were previously known, as well as 14 new putative serine proteases. Among the McSPs, McSP30, 32, 33, 38, 54, 57 and 59, are likely inactive serine proteases since they lack a GXSXG motif and/or a complete catalytic triad composed of H^{57}, D^{102}, and S^{195} (Krem et al., 2000). Each of the PM associated McSPs (except McSP38 and 48) contained five conserved cysteines. Wang et al. (1997) reported that C^{191} forms a disulphide bridge with C^{220} to stabilize the substrate binding pocket.

Serine protease specificity can be inferred from the amino acid at position 189 which is at the base of the S1 binding pocket (Hegedus et al., 2003). The presence of an aspartic acid, serine

or a glycine in this location suggests trypsin, chymotrypsin or elastase-like activity, respectively (Hegedus et al., 2003; Erlandson et al., 2010). Eight trypsins (with D^{189}), fifteen chymotrypsins (with S^{189}) and nine elastases (G^{189}) are associated with the *M. configurata* PM. Although McSP48, which had N^{189}, could not be categorized based on the specificity amino acid, Erlandson et al. (2010) reported that it clustered with chymotrypsins based on a phylogenetic analysis.

Association of serine proteases with the PM has been reported previously (Eguchi et al., 1982; Ferreira et al., 1994). More recently, particular serine proteases including six trypsin and four chymotrypsin enzymes have been identified from the *H. armigera* PM (Campbell et al., 2008). Campbell et al. (2008) suggested that the association of serine proteases with the PM may simply reflect transit across it to their final destination in the gut lumen. However, Terra (2001) has proposed that the PM may serve as an anchor for digestive enzymes and that the interaction of these enzymes with the PM may be of a more integral nature. The current study is only the second demonstration of a large diversity of serine proteases associated with insect PM and may support the proposal of Terra (2001).

4.4.2.2.2 Exopeptidases

Exopeptidases are the major proteases in the ectoperitrophic space; they hydrolyse peptides produced from the initial digestion of proteins by endopeptidases (Billingsley, 1990; Lee and Anstee, 1995). They are membrane bound metalloenzymes that remove a single amino acid from the amino (aminopeptidases) or carboxy terminus (carboxypeptidases) of peptides (Ferreira et al., 1994; Terra and Ferreira, 1994). In *M. configurata*, six aminopeptidases and five carboxypeptidases are associated with the PM.

Insect digestive aminopeptidases belong to the family aminopeptidase N (APN) (Hooper, 1994). All *M. configurata* aminopeptidases have the typical features of this enzyme family such as a signal peptide, a GPI anchor sequence at the carboxy terminus, and the zinc binding motif "HEXXH" (Jongeneel et al., 1989; Hooper, 1994). A serine-threonine-proline rich region upstream of the GPI anchor sequence is also present in McAPN4, 6 and 7 as reported for other several APNs (Piggot and Ellar, 2007; Angelucci et al., 2008). Serine-threonine-proline rich regions and GPI were reported to be putative sites for *Bacillus thuringiensis* Cry1Ac toxin binding (Liao et al., 1996; Oltean et al., 1999). APNs are expressed primarily in midgut (Piggot

and Ellar, 2007; Angelucci et al., 2008) and association of aminopeptidases with PM has been shown previously (Terra et al., 1979; Walker et al., 1980).

Insect digestive carboxypeptidases belong to the peptidase family M14 (Wang et al., 2004b) and all *M. configurata* carboxypeptidases, except McCPA5, have the basic features of this family of enzymes such as a signal peptide, a proenzyme (with approximately 90 amino acids), and a mature enzyme (of approximately 310 amino acids) that contains a zinc binding motif (H^{69}, E^{72} and H^{196}) (Edwards et al., 2000; Yan et al. 2002). A pair of conserved cysteines (C^{138} and C^{161}) that form a disulphide bond important in maintaining the tertiary structure of the enzyme (Yan et al. 2002; Wang et al., 2004b) is present in McCPB1, McCPB4 and McCPA6. Several arginines are present at the amino terminus of McCPs that corresponds to the region of *H. armigera* CPA which was cleaved by trypsins causing the enzyme to be released from the membrane (Bayes et al., 2005). Wang et al. (2004b) reported that *T. ni* carboxypeptidases have an arginine or a lysine in the same region.

Insect carboxypeptidases have been classified as CPA (specific to hydrophobic, aromatic and aliphatic amino acids), CPB (basic amino acids) and CPC (broad specificity) (Titani et al., 1975; Rees et al., 1983; Rawlings, 1998). The specificity amino acid (position 255) in CPAs is noncharged, but charged in CPBs (Bown et al., 1998; Wang et al., 2004b). Both CPAs and CPBs are associated with the *M. configurata* PM; however, CPA, CPC and glutamate type carboxypeptidases were associated with *H. armigera* PM (Campbell et al., 2008). The midgut is the major site for expression of carboxypeptidase genes (Ramos et al., 1993; Bown et al., 1998) and association of carboxypeptidases with the PM has been shown previously (Ferreira et al., 1994; Bown et al., 1998).

4.4.2.2.3 Insect intestinal lipases

IILs are esterases that aid digestion of dietary fats in the midgut. Eight McIILs are associated with the *M. configurata* PM. IILs have a conserved catalytic triad comprised of serine, histidine and aspartic acid (Brady et al., 1990), in which the serine is part of the conserved "GXSXG" active site motif (Hide et al., 1992). McIIL1, 5, 6, 7, and 9 all have this active site motif; however, McIIL3 and McIIL4 have a "GXGXG" motif in the corresponding region, suggesting they may not be active (Figure 4.5). Similarly, several IILs from *E. postvittana* had a glycine or glutamic acid instead of serine in the active site, suggesting they are not active

(Simpson et al., 2007). Four lipases, of which two were proposed to be inactive, have also been reported from *H. armigera* PM (Campbell et al., 2008). All IILs, except McIIL9, have a conserved arrangement of cysteines, $C^1X_{4-9}C^2X_{22-25}C^3X_{10}C^4$, near the carboxy terminus of the proteins which may be involved in binding of IILs to the PM.

4.4.2.2.4 β-1, 3-Glucanase

Midgut β1, 3GLUs are involved in digestion and/or immunity (Genta et al., 2003; 2009; Pauchet et al., 2009a) and are secreted into the lumen of the larval gut to hydrolyze β-1, 3-glycosidic bonds in β-1, 3-glucans (Genta et al., 2009; Pauchet et al., 2009a). One β1, 3GLU, Mcβ1, 3GLU, is associated with the *M. configurata* PM. A glycosyl hydrolase family 16 domain is present at the carboxy terminus of Mcβ1, 3GLU, as a common feature of β1, 3GLUs (Ma and Kanost, 2000; Yu et al., 2002). *Mcβ1, 3GLU* is expressed in foregut, midgut, hindgut, Malpighian tubules, tracheae, integument and fat body as was shown for pattern recognition carbohydrate binding proteins involved in insect innate immunity (Pauchet et al., 2009a; Bragatto et al., 2010). Mcβ1, 3GLU is the second β1, 3GLU shown to be associated with the PM after Haβ1, 3GLU (Campbell et al., 2008).

4.4.2.2.5 α-Amylase

α-Amylase catalyzes the hydrolysis of the α-(1, 4) glycosidic linkage of large polysaccharides such as starch and glycogen (Janecek, 1997). Midgut α-amylases have been identified from coleopterans (Grossi de Sa and Chrispeels, 1997; Strobl et al., 1998) and lepidopterans (Pauchet et al., 2008; Pytelkova et al., 2009). In *M. configurata*, one α-amylase, McAMY, is associated with the PM. Association of an α-amylase with lepidopteran PM has been demonstrated previously (Ferreira et al., 1994). Three to seven disulphide bonds were reported in different insect α-amylases (Strobl et al., 1998; Pelegrini et al., 2006; Pytelkova et al., 2009). McAMY has 10 cysteines, suggesting the formation of 5 putative disulphide bridges, which may be involved in binding the enzyme into the PM or stabilizing the active site.

4.4.2.3 Other Enzymes

4.4.2.3.1 Alkaline phosphatases

Alkaline phosphatases (ALPs) remove phosphate groups from organic molecules such as nucleotides, proteins, and alkaloids (Perera et al., 2009). One ALP, McALP1, is associated with

the *M. configurata* PM. McALP1 is predicted to have several *N*-glycosylated asparagines, which have been shown to be important binding sites for *B. thuringiensis* Cry toxins (Jurat-Fuentes and Adang, 2004; Ning et al., 2010), and a GPI anchor linked to S^{445}, as was reported for ALPs from *Heliothis virescens* (Jurat-Fuentes and Adang, 2004) and *Manduca sexta* (Chen et al., 2005). *McALP1* expression is restricted to the midgut and ALPs have been reported to be present in the midgut (Chen et al., 2005; Fernandez et al., 2006) or specifically PM (Campbell et al., 2008) from other insects.

4.4.2.3.2 dsRNase

dsRNase is an endoribonuclease that cleaves double-stranded RNA. An alkaline nonspecific nuclease from the midgut of silkworm, BmdsRNase, had RNase activity and degraded DNA to a lesser degree (Arimatsu et al., 2007a). An ortholog of BmdsRNase, McdsRNase, is associated with the *M. configurata* PM. Based on the domain features of BmdsRNase, McdsRNase is predicted to undergo amino terminal processing to release the mature active enzyme outside the cell; this may protect the cellular nucleic acids from being degraded by this dsRNase. *McdsRNase* expression is restricted to the midgut as was also shown for *BmdsRNase* (Arimatsu et al., 2007b). Midgut specific expression and association with the PM suggests that McdsRNase is involved in the digestion of nucleic acids in the diet. In other studies, dsRNase activity was highest during feeding, but decreased during molting and pupation (Koga et al., 1969). dsRNases may also protect the midgut epithelium against viral pathogens by degrading their genetic material (Furusawa et al., 1993).

4.4.2.3.3 Astacin

Astacins are metalloproteinases that are either secreted from cells or anchored to the plasma membrane (Bond and Beynon, 1995). In *M. configurata*, one astacin protein, McAST, is associated with the PM. McAST has a putative tryptic cleavage site, in 18-RSRS-23, suggesting it may be cleaved for activation. McAST also has a typical zinc binding site “HEXXH”, a common feature of metalloproteinases (Gomis-Rüth et al., 1993), and four conserved cysteines. Two of the cysteines, C^{148} and C^{174}, correspond to C^{42} and C^{198} in *Astacus astacus* astacin, which have been shown to form a disulphide bridge (Bode et al., 1992; Stöcker et al., 1993). Thus, C^{148} and C^{174} and the remaining two cysteines C^{124} and C^{284} may form two disulphide bridges in

McAST. Similar cysteine arrangements, $C^1X_{21}C^2X_{19}C^3X_{113}C^4$ in *Astacus* astacin and $C^1X_{22}C^2X_{25}C^3X_{109}C^4$ in McAST, support this prediction. These putative disulphide bridges may be involved in the binding of the enzyme into the PM or structural stabilization of the enzyme.

McAST expression is specific to midgut as was reported for *GmZmp* in *G. m. morsitans* (Yan et al., 2002). Other midgut astacins have been also reported from *Melolontha melolontha* (Wagner et al., 2002), *S. frugiperda* (Ferreira et al., 2007), *L. longipalpis* (Jochim et al., 2008) and *Diabrotica virgifera virgifera* (Kaiser-Alexnat, 2009). Astacins are typically involved in digestion (Vogt et al., 1989; Yan et al., 2002) and McAST midgut specific expression and its PM localization also suggest a digestive role.

4.4.2.3.4 Pantetheinase

Pantetheinase hydrolyzes pantetheine, an intermediate metabolite of coenzyme A, into pantothenic acid (vitamin B5) and cysteamine, a potent antioxidant (Pitari et al., 2000). A putative pantetheinase, McPAN, is associated with the *M. configurata* PM. Although McPAN lacks sequences at the amino terminus, the full length orthologs from *N. vitripennis*, *T. castaneum* and *C. quinquefasciatus* have signal peptides, suggesting McPAN is secreted for PM association. Interestingly, McPAN has 20 putative *O*-glycosylated amino acids, suggesting the presence of a mucin-like domain. In contrast, the brain pantetheinase in *Drosophila melanogaster* was reported to be *N*-glycosylated (Koles et al., 2007).

McPAN is expressed in hindgut and Malpighian tubules in addition to midgut (Figure 4.3). Several membrane bound pantetheinases were reported from midgut (Ferreira et al., 2007), silk gland (Zhang et al., 2006) and brain (Koles et al., 2007); however, association of a pantetheinase with the PM has not previously been shown and its function in insects is not known.

4.4.3 Other Proteins

4.4.3.1 HMG176-like proteins

A protein, McMG176, similar to HMG176 from *H. armigera* (Dong et al., 2007), is associated with the *M. configurata* PM. Like HMG176 (Wang et al., 2007), McMG176 has a transmembrane domain at the carboxy terminus, suggesting that it is cleaved prior to being incorporated into the PM. However, *McMG176* is expressed only in the midgut of feeding

larvae; whereas *HMG176* is expressed in the midgut and fat body primarily during molting (Dong et al., 2007; Wang et al., 2007). Although McMG176 is associated with the PM, HMG176 is localized to the basal lamina of the midgut and the fat body (Wang et al., 2007).

4.4.3.2 Lipocalin

Polymeric lipocalins, or polycalins are secreted proteins with an affinity for binding small hydrophobic ligands in midgut (Pandian et al., 2008; Pauchet et al., 2008). A single polycalin with seven lipocalin domains, McPOL, is associated with the *M. configurata* PM. A polycalin was also reported from *H. armigera* PM (Campbell et al., 2008) and *B. mori* midgut (Mauchamp et al., 2006), suggesting they may be a common feature of lepidopteran midgut and/or PM. Expression of *McPOL* is restricted to midgut and midgut polycalins have been implicated in binding of the plant pigment chlorophyllid A (Mauchamp et al., 2006) as well as to *B. thuringiensis* Cry toxins (Angelucci et al., 2008).

4.4.3.3 REPAT

Four midgut genes were found to be upregulated in response to *B. thuringiensis* and baculovirus infections in *S. exigua* larvae and were denoted "response to pathogen-*REPAT*" genes (Herrero et al., 2007). An ortholog of SeREPAT2, McREPAT2, is associated with the *M. configurata* PM, which is the first report of such an association. *McREPAT2* is expressed only in the midgut; however, *SeREPAT* genes are expressed also in hindgut (Herrero et al., 2007). Upregulation of *REPAT* genes in pathogenic infections suggests a defensive role against pathogens (Herrero et al., 2007). However, the reason for the association of McREPAT2 with PM in healthy larvae is not known.

4.4.3.4 Serpin

Serpins are irreversible inhibitors of serine proteases that regulate proteolytic activities (Carrell and Boswell, 1986). All serpins possess a conserved domain of 370-390 amino acids and a 30-40 amino acid variable carboxy terminal domain (Molnar et al., 2001). The variable domain forms the inhibitory reactive centre loop that inserts into the protease active site (Ye and Goldsmith, 2001). McSerpin1A is an isoform derived from the McSerpin1 mRNA through differential splicing of the transcript (Hegedus et al., 2008). The *McSerpin1A* variant was

expressed in the foregut, midgut, hemocytes, fat body and tracheae (Hegedus et al., 2008) and the protein was found to be associated with the PM (Chamankhah et al., 2003).

4.4.3.5 CLECT

CLECTs are carbohydrate binding proteins in animals and the C type designation is due to their requirement for calcium for binding. One such lectin, McCLECT, is associated with the *M. configurata* PM. McCLECT and other lepidopteran CLECTs (Pauchet et al., 2010) have two CRDs each containing several Ca^{2+} binding sites and four to six conserved cysteines forming two to three putative disulphide bridges. *McCLECT* is expressed in foregut, hindgut, Malpighian tubules, tracheae, integument and fat body in addition to midgut, which is similar to the expression patterns shown for other CLECT genes (Tanji et al., 2006; Takase et al., 2009; Pauchet et al., 2010). Although several midgut lectins have been shown to be associated with the PM (Peters et al., 1983; Lehane and Msangi, 1991), McCLECT is the first CLECT shown to be associated with the PM, and CRDs are likely responsible for this association.

4.4.3.6 Lsti99 and Lsti201-like proteins

Two proteins (363-100-1 and 530-224-1) that are similar to *L. sticticalis* PM proteins, Lsti99 and Lsti201 (Yin et al., 2010), are associated with the *M. configurata* PM. Alignment of 363-100-1 with Lsti99 and Lsti201 revealed several conserved cysteines; however, they do not conform to any known CBD consensus. *530-224-1* is expressed only in midgut, whereas *363-100-1* is also expressed in hindgut, Malpighian tubules and tracheae. 363-100-1 and 530-224-1 may be localized to other tissues as was the case for Lsti99 and Lsti201 PM proteins which were also detected in head, hemolymph and integument in *L. sticticalis* larvae (Yin et al., 2010).

4.4.3.7 Soybean lectin

Soybean lectin, GmLe1, is a member of the legume-type lectin family (Loris et al., 1998). GmLe1 undoubtedly originated from the artificial diet fed to *M. configurata* and its association with *M. configurata* PM has been previously shown (Shi et al., 2004).

4.4.4 Proteins without Orthologs

Three novel proteins 357-94-1, 530-247-1 and 357-80-1 are associated with the *M. configurata* PM. These proteins are all predicted to have a signal peptide, suggesting they are secreted. 357-94-1 and 530-247-1 are closely related (identity value: 88%) and have eight conserved cysteines which does not conform to any CBD cysteine arrangement. 357-80-1 has various *O*-glycosylated threonines or serines interspersed with prolines, suggesting the presence of a MD; however, no cysteine is present in this protein. Genes encoding 357-94-1 and 530-247-1 are expressed only in the midgut, while the gene encoding 357-80-1 is also expressed in Malpighian tubules.

In conclusion, 82 PM proteins were classified into three groups: peritrophins, enzymes and other proteins. The six peritrophins included IIMs and nonmucins and the associated enzymes included 2 CDAs, 54 putative digestive enzymes (serine proteases, aminopeptidases, carboxypeptidases, lipases, β-1, 3-glucanase and α-amylase) and 4 other enzymes (Alkaline phosphatase, dsRNase, astacin, and pantetheinase). The third group contained 13 known proteins: polycalin, REPAT, serpin, CLECT, MG176, Lsti99-like and Lsti201-like proteins. Three proteins had no identified orthologs in the NCBI database. Expression of the majority of the genes encoding PM proteins was specific to the midgut. This study is the first report of association of dsRNase, astacin, pantetheinase, REPAT, CLECT, and MG176, as well as three PM proteins without orthologs, with the PM. In addition, expression of some peritrophin genes (e.g., *McPM1, McPPAD1* and *McCBD3P*) is not restricted to tissues synthesizing PM, although expression of McIIM genes is restricted to midgut.

In the following chapter, a detailed characterization of select PM proteins including IIMs, CDAs, IILs and serine proteases will be presented. Tissue specific gene expression patterns and protein localization studies in different developmental stages and feeding conditions are summarized. In addition, enzymatic activity experiments for CDAs and serine proteases are presented.

5. CHARACTERIZATION OF SELECTED PERITROPHIC MATRIX PROTEINS[3]

5.1 Introduction

Proteins are the main component of the peritrophic matrix (PM) (Peters, 1992) with peritrophins and enzymes being the most abundant. The peritrophins comprise solely structural proteins having characteristic chitin binding domains (CBDs) and may also contain mucin domains (MDs) as in the case of insect intestinal mucins (IIMs). In vertebrates, mucins protect epithelial cells from pathogens, physical damage and proteases as well as lubricate the passage of food (Van Klinken et al., 1995). IIMs were proposed to play similar roles (Wang and Granados, 1997a). They have been found in many species (Wang and Granados, 1997b; Shen et al., 1999; Venancio et al., 2009) and are most often associated with the PM, but occasionally with the microvillar membrane (Shen et al., 1999).

Enzymes associated with the PM are primarily involved in digestive processes and chitin modification. Insect digestive protease composition is largely reflective of the lumen pH. The midgut of lepidopteran larvae is alkaline and serine proteases are the primary enzymes used to digest proteins (Christeller et al., 1992). Insect intestinal lipases (IILs) form another large enzyme family and digest dietary fats (Pistillo et al., 1998). An enzyme involved in the modification of chitin is chitin deacetylase (CDA) and its association with the PM was first demonstrated in *Trichoplusia ni* larvae (Guo et al., 2005). CDAs catalyze the conversion of chitin into chitosan; however, neither the presence of chitosan nor CDA activity have been detected in any insect system.

The current chapter focuses on the further characterization of select *Mamestra configurata* PM proteins including IIMs, CDAs, IILs and serine proteases. IIMs were selected as they are one of the two classes of peritrophins. Expression and localization of McIIMs were examined for feeding, molting and starved larvae. McCDA1 was investigated as it was a predominant PM protein and may be involved in chitin modification. The midgut was shown to have CDA activity and McCDA1 was active. Two types of digestive enzymes associated with the PM, serine proteases and IILs, which are the major protein and fat degrading enzymes, respectively, were also characterized. These studies included examination of gene expression patterns, protein localization and serine protease activity.

[3] The contents of this chapter are published in *Insect Molecular Biology*, **17:** 573-585, 2008 and **19:**163-175, 2010.

5.2 Material and Methods

5.2.1 Rapid Amplification of cDNA Ends (RACE)

5' RACE was used to obtain the entire coding regions of selected cDNAs. Gene specific primer 1 (GSP1) was used to convert target mRNA into cDNA and gene specific primer 2 (GSP2) amplified the 5' ends from the cDNA (GSP2) (Table 5.1). The 5' ends of *M. configurata* insect intestinal mucin 4 (McIIM4) and *M. configurata* insect intestinal lipase 2 (McIIL2) cDNAs were obtained using the GeneRacer™ Advanced RACE Kit (Invitrogen, Carlsbad, CA, USA), whereas the 5' end of *M. configurata* insect intestinal mucin 2 (McIIM2), *M. configurata* insect intestinal mucin 3 (McIIM3), *M. configurata* chitin deacetylase 1 (McCDA1), *M. configurata* insect intestinal lipase 1 (McIIL1) and *M. configurata* insect intestinal lipase 3 (McIIL3) cDNAs were obtained using the 5' RACE System (Invitrogen).

Table 5.1 Primers used for 5' RACE analysis of *M. configurata* McCDA1, McIIL1-3 and McIIM2-4.

Primer[1]	**Sequence**
McCDA1 GSP1	CCAAGCCATATTCTTTCATG
McCDA1 GSP2	TGACTGATGGAGTGTAGAGC
McIIL1 GSP1	TCCTCGCTACAGTTTCCAGC
McIIL1 GSP2	CGCCCATGTAGAAATATGACC
McIIL2 GSP1	GAGTTGGTCTCCAAGTAGAAAATACC
McIIL2 GSP2	TTAGGCACAAGCCCACCCATCCTAAGG
McIIL3 GSP1	TTCTGCACCTATCAGTGGTAC
McIIL3 GSP2	GTTCTCCAGTCAACCTCTATG
McIIM2 GSP1	TCCGTACAATCTACGTTGTG
McIIM2 GSP2	AGGACACGTGTATTCAATGG
McIIM3 GSP1	GTTGTCTCCGTGTAAGTTCC
McIIM3 GSP2	AGTTTGAGGAGTTAGGGTCGG
McIIM4 GSP1	ATGTTCTATCAACTCAAGCATATTAGCTT
McIIM4 GSP2	TCCCTCAGAGCAGATGTTAAGACGA

[1] **Abbreviations:** CDA, chitin deacetylase; GSP, gene specific primer; IIL, insect intestinal lipase; IIM, insect intestinal mucin; Mc, *M. configurata*.

5.2.2 RNA Extraction and Analysis of Gene Expression

Expression of target genes was examined using reverse transcription-PCR (RT-PCR) or northern blot analyses with total RNA extracted from various tissues as desribed in Chapter 2. Expression of genes encoding McIIMs, McIILs and serine proteases was examined at four time points, including three stages of feeding (early, mid, and late feeding) and one stage of molting (Phase II) larvae. *McCDA1* expression was examined in two additional molting stages (Phases I and III) in addition to the four time points above. The molting phases were defined in Chapter 2. Primers used in RT-PCR analysis are shown in Table 5.2.

Table 5.2 Primers used in expression analysis of *McCDA1*, *McIIL*, *McIIM* and *McSP* genes by RT-PCR.

Primer[1]	**Sequence**
McACT Fp	GATCGTGCGTGACATCAAGG
McACT Rp	CACACTTCATGATCGAGTTG
McCDA1 Fp	CGAGGCTTTCTTGTCAGCTTTC
McCDA1 Rp	TTGTCCAAGAGGGTTGCCGAG
McIIL1 Fp	GAGGCGATGTGGACTTCTAC
McIIL1 Rp	CGGAGATGGGTTTGTTCTTA
McIIL2 Fp	ATGGCAGAGTCTCTGCAATC
McIIL2 Rp	TCAAGTATCTTCAGCCGCG
McIIL3 Fp	TACGTCGAAGTCATCCACAC
McIIL3 Rp	GTAAAGTACAATAGAAGTGAC
McIIM1 Fp	AAACCCTTCTACTCGTGACG
McIIM1 Rp	GGTGGGTGCAGCAGTTTCAC
McIIM2 Fp	CTACCACAACGACTCCTGCA
McIIM2 Rp	GAACGCGTTGCAGTCACTCTC
McIIM3 Fp	ACGACTCTTTGTGATGCTTC
McIIM3 Rp	CTCGTTATAAGCTGTTAGTC
McIIM4 Fp	CAACATCAGCACCTGACTGC
McIIM4 Rp	CTCAGAGCAGATGTTAAGACGA
McSP1 Fp	AAGCTTATGCGTTTGAGTTGAT
McSP1 Rp	GCGTCACGTCGAGCTTAGGGTTAT

Table 5.2 (continued)

Primer[1]	Sequence
McSP2 Fp	GTTCAGCAAGGCACCATCATC
McSP2 Rp	TCACCGAAGTACAGAGGACCAC
McSP25 Fp	CCCAGGTAGCCAGGTCATT
McSP25 Rp	ATTCGGGCTTGTTTTCATTG
McSP29 Fp	CTCACCAATGACGGTGGTAGC
McSP29 Rp	CCGATCAAGATATGCCTGCTGC
McTUB Fp	AGCGTACCATCCAGTTCGTG
McTUB Rp	TGGCGTACATGAGGTCGAAC

[1]**Abbreviations:** ACT, actin; CDA: chitin deacetylase; Fp, forward primer; IIL: insect intestinal lipase; IIM: insect intestinal mucin; Mc, *M. configurata*; Rp, reverse primer; SP: serine protease, TUB, tubulin.

5.2.3 Expression of Recombinant Proteins in Escherichia coli

Recombinant McIIM2, McIIM3 and McCDA1 were produced using the One-Step RT-PCR kit (Invitrogen) with the primers (Table 5.3) as described in Chapter 2.

Table 5.3 Primers used to amplify open reading frames of *M. configurata* McCDA1, McIIM2 and McIIM3 to generate constructs for expression of recombinant proteins in *E. coli*.

Primer[1]	Sequence
McCDA1 Fp	GGAATTCCATATGGATTCTTCTTCAAAAGAGGAAGC
McCDA1 Rp	CCGCTCGAGTTAAAGAGGGTTGCCGAGCCATGGGTAGACACG
McIIM2 Fp	GGAATTCCATATGCAAAACACAATGCTCGTAGATAAG
McIIM2 Rp	CCGCTCGAGTCATCGCATCATTTGTCTGGAGTCCAA
McIIM3 Fp	GGAATTCCATATGGTTGTAAACTGCACGGCCACCGGA
McIIM3 Rp	CCGCTCGAGTCAATTAGTACAAGTATACGTTGCTGT

[1]**Abbreviations:** CDA: chitin deacetylase; Fp, forward primer; IIM: insect intestinal mucin; Mc, *M. configurata*; Rp, reverse primer.

5.2.4 Protein Extraction and Separation, and Localization of Target Proteins

Western blot analyses were used to determine the localization of proteins. The proteins associated with the PM and midgut samples were extracted as strongly or loosely associated fractions and resolved using one dimensional (1D) or two dimensional (2D) SDS-PAGE as

described in Chapter 2. Anti-McCDA1, McIIM3 and McIIM4 antisera were generated using conjugated peptides (EzBioLab, Carmel, IN, USA) or (Sheldon Biotechnology Centre, McGill University, Montreal, QB, Canada) (Table 5.4) while the anti-McIIM2 antiserum was generated against a purified recombinant protein.

Table 5.4 Sequences of the conjugated peptides used to generate antisera specific for *M. configurata* McCDA1, McIIM3 and McIIM4.

Peptide name[1]	**Company**	**Sequence**
McCDA1 P1	Sheldon Biotechnology Centre	EECTPETVCELPNCRCSSTN
McCDA1 P2	Sheldon Biotechnology Centre	RRFNPRACVARSCGPLNSGH
McIIM3 P1	Sheldon Biotechnology Centre	SDGTFTQYNYTCPSTSSFNP
McIIM3 P2	Sheldon Biotechnology Centre	DPADATCKNYTLCVYVSSTN
McIIM3 P1	EzBioLab	TYTCPATTTNTTTSVCTADG
McIIM3 P2	EzBioLab	YTCPATTLYNPNTTLCDASYC
McIIM4 P1	EzBioLab	EEVPAPADPELPGTTTNSPPC
McIIM4 P2	EzBioLab	DGMLIFVPHSFNNKANMLEC

[1]**Abbreviations:** CDA: chitin deacetylase; IIM: insect intestinal mucin; Mc, *M. configurata*; P1, peptide 1; P2, peptide 2.

5.2.5 Phylogenetic Analysis of Chitin Deacetylases

A set of 23 CDAs comprised of 11 from *Tribolium castaneum*, 10 from lepidopterans and 2 from dipterans were obtained from the NCBI database and subjected to phylogenetic analysis. Dendrograms arising from phylogenetic analysis were constructed according to the neighbor-joining method and confidence values for the branches determined using bootstrap analysis where 100 trees were generated from randomly resampled data generated by CLUSTALW.16 as provided in PHYLIP version 3.5C (www.cbr.nrc.ca/cgibin/WebPhylip/index.html). The CLUSTALW alignment was performed using the Blosum scoring matrix with the following gap penalties; opening gap (10), end gap (10), extending gap (0.05) and separation gap (0.05).

5.2.6 Demonstration of Chitin Deacetylase Activity

CDA activity was assessed using the method of Trudel and Asselin (1990), with the following modifications. Samples consisting of loosely associated midgut and PM proteins or recombinant McCDA1 or McPPAD3 (negative control) were mixed with nonreducing sample

buffer (without β-mercaptoethanol) and proteins were separated on 12% PAGE gels containing 0.5% SDS. Proteins were transferred to a 7.5% PAGE-HEPES-KOH (pH 7.0) gel containing 0.1% (w/v) reacetylated glycol chitosan and 0.1% (w/v) SDS by electroblotting at 60V for 30 min. The glycol chitin gel was soaked in 100 ml of 50 mM HEPES-KOH (pH 7.0) and 1% (v/v) Triton X-100 for 75 min at 37^{o}C to displace SDS, allowing for enzyme activation and hydrolysis of substrate. The gel was then incubated in 20 ml of 0.01% (w/v) fluosrecent brightener 28 (FB28) (Sigma, St Louis, MO, USA) in 500 mM Tris-HCl (pH 8.8) for 5 min. The FB28 solution was removed and the gel destained in distilled water for 48 h at room temperature with frequent changes. Deacetylated glycol chitin appeared as highly fluorescent bands on a UV transilluminator. The original separating gel was stained with Coomassie blue to identify the lanes showing zones of fluorescence in the glycol chitin gel. The proteins from a separate gel were subjected to western blot analysis as described above.

5.2.7 Chitosan Test

To detect whether the PM contains chitosan, a freshly dissected PM and a PM that had been stripped of strongly associated proteins (as described in Chapter 2) were used according to Peters (1992). As negative controls, PM-free midguts and starch were used. Samples were acidified to pH 2.5 with sulphuric acid, followed by 3% acetic acid to wash out the sulphuric acid. Samples were then treated with potassium iodide and checked for production of a reddish-violet colour as a result of the reaction of potassium iodide with chitosan.

5.2.8 Protease Assays

Serine protease isoforms were resolved using the PAGE zymographic method of Schmidt et al. (1988) as modified by Hegedus et al. (2003). The whole midgut containing the PM and food bolus was isolated from feeding or 24 h starved 5th instar larvae, or from larvae molting from the 5th to the 6th instar, and suspended in 200 µl ddH_2O. Two midguts were pooled and ground with a pestle followed by centrifugation at 20,800 g for 1 min. One microliter of the supernatant was used for electrophoresis and isoform detection. Total midgut serine protease activity was quantified using azoalbumin as a general protein substrate according to Hegedus et al. (2003). A 5 microliter aliquot of each midgut homogenate prepared as described above was used for each assay.

5.3 Results

5.3.1 One Dimensional Separation of PM Proteins in Feeding, Molting and Starved Larvae

1D SDS-PAGE analysis revealed that the protein profiles differed in the PMs of starved, molting and feeding larvae, with approximately 18, 30 and 24 bands being detected, respectively (Figure 5.1). Many of the protein bands were present under all experimental conditions. The PM protein profile from starved larvae had seven bands that were also seen in the profile from molting larvae; however, these seven bands were not apparent in the profile from feeding larvae. The PM protein profile from molting larvae had five unique bands and eight bands that were also present in the profile from the PM of feeding larvae. Five bands were unique to the protein profile of the PM from feeding larvae. Eleven bands representing proteins of greater than 25 kDa were common to the profiles of all PMs, as was an especially prominent band of approximately 28 kDa (see below). In addition, the profiles of all PMs had 2-6 lower molecular weight bands (≤25 kDa). The profiles of the major protein bands were consistent over several experiments with this cohort of insects, though the profiles of the minor bands could vary.

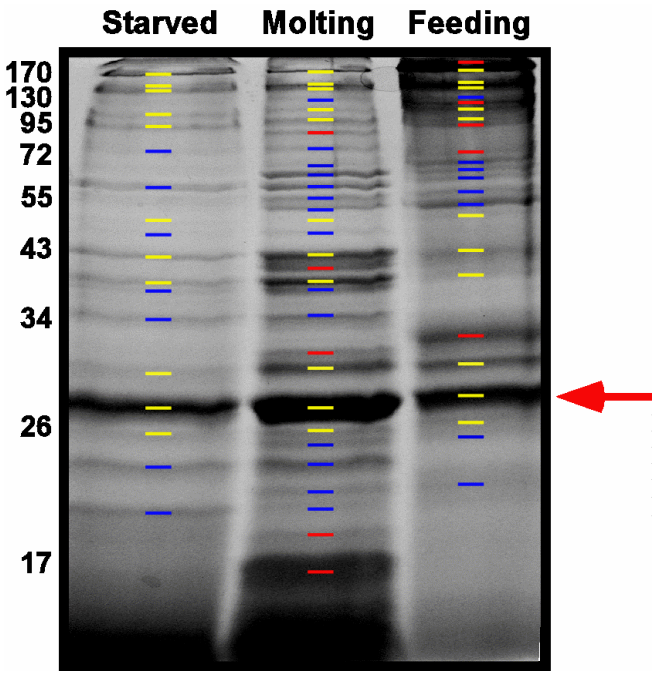


Figure 5.1 Separation of strongly associated peritrophic matrix proteins using one dimensional SDS-PAGE denaturing gel electrophoresis. Proteins unique to feeding, starved or molting larvae (red lines), common to at least two of these (blue lines) or all (yellow lines) are indicated on the bands in each lane. The red arrow indicates the prominent 28 kDa band. The sizes of the molecular weight markers (kDa) are shown in the left hand margin.

5.3.2 Insect Intestinal Mucins (McIIM2-4)

5.3.2.1 Identification

McIIM2 and McIIM4 had already been identified through liquid chromatography tandem-mass spectrometry (LC-MS/MS) in Chapter 4. 5' RACE analysis revealed that the McIIM2 cDNA was 1083 bp, 6 bp longer than the sequence obtained from the cDNA library; however, the open reading frame (ORF) was the same. McIIM2 had numerous serines or threonines that were predicted to be *O*-glycosylated in the region between T^{92} and T^{235}, but had no putative *N*-linked glycosylation sites (Figure 5.2). Overall, 92% of the threonines and 50% of the serines were predicted to be *O*-glycosylated in McIIM2. 5' RACE analysis revealed six different alleles for *McIIM4* (McIIM4.1 to McIIM4.6) that ranged from 2789 to 2861 bp and encoded proteins of 627 to 651 amino acids (Figure 5.2). The main variation among the derived proteins was in a region between amino acids 342 and 366. The largest protein (McIIM4.1) had a sequence "DNSTTPVYEITTTPVSDNSTTPVS" in this region, whereas McIIM4.2-6 had partial or complete deletions of this sequence (Figure 5.2). McIIM4 was predicted to have a molecular weight of 65.5-68 kDa after removal of the signal peptide. The proteins had putative *O*-linked glycosylation sites between T^{198} and T^{214} as well as T^{304} and T^{555} and had 12 *N*-linked glycosylation sites at $N^{121, 316, 325, 343, 359, 367, 383, 391, 415, 439, 463, 520}$. In total, 81% of the threonines and 61% of the serines were predicted to be *O*-glycosylated in McIIM4.

A search of the EST database for proteins with CBDs identified an additional IIM and 5' RACE was used to obtain the entire coding region of this cDNA. The cDNA was 796 bp, and encoded a 211 amino acid protein with a 21 amino acid signal peptide (Figure 5.2). The protein, denoted McIIM3, had a predicted molecular weight of 20 kDa without the signal peptide. The protein had putative *O*-linked glycosylation sites between T^{72} and T^{95} as well as T^{204} and T^{210} and had seven putative *N*-linked glycosylation sites at $N^{24, 43, 107, 118, 137, 146, 187}$. Overall, 34% of the threonines but none of the serines were predicted to be *O*-glycosylated in McIIM3.

5.3.2.2 Structural features and organization

Each McIIM had a unique structure. McIIM2 had one MD separating two CBDs, while McIIM3 had three CBDs and two small MDs (Figure 5.2). McIIM3 was unique in that the first MD overlapped with the first CBD. Likewise, the second MD overlapped with the carboxy

terminal CBD resulting in a hybrid domain having the characteristics of both a CBD and a MD. McIIM4 had four CBDs and two MDs.

McIIM2

```
MKGIIILLFV VYAAVGRAQN TMLVDKCPEE QETDWTIELL LRHDDCNKFY KCTFGKPIEY TCPADLWFDL  70
DTQQCDWRHN VDCTDRYVPG ETPTTEPTTA APTTTEPTTL PTTTPTTTTT TTTTPAPTTT TSTTTTTTPA 140
PTTTTSTTTT TTPAPTTTTT TTTTPAPTTT TTTTTPAPTT TPTTTTTTTT PAPTTTPTTT TTTTTPAPTT 210
TPTTTTTTTT PAPTTTTSTT AAPITFLPNG CPTNPHIHWL LPHESDCNAF YYCVWGQLVL RQCPATLHFN 280
RVIQVCDWPW DAGCPVSLNK HLDSRQMMR                                             309
```

McIIM3

```
MAKQIIIIVF MLSICAWQAH AVVNCTATGA GRQADPADAT CKNYTLCVYV SSTNTYVSYN YVCPTTSLFS  70
PTLRQCTTTY TCPATTTNTT TSVCTADGFI ADPNSSNCSS YIECVNINGT YTETTYTCPA TTLYNPNTTL 140
CDASYNCTST TAFTCTTAGR FANTADTTCQ TYFYCVLLSD GTFTQYNYTC PSTSSFNPAS SLCTATYTCT 210
N                                                                           211
```

McIIM4

```
MYKTLFFLTA LAIVQARLNE DAVVTNVRLN KVPVQIANAD DDCETLDNGC PVDFTIHKLV PHEEYCHLFY  70
YCDKGELVLS SCPEPLYFDP KAQVCVWSWA TDCVNNGPYT YPTTAAPEVE NSTAPGTIDI GEVLDNGCPS 140
DIHIHHHLPH EECEKFYQCN FGQKVERDCA PGTVFHFEIQ VCDWPRNVPR CAGSAGATAR PQTTPEASSE 210
EIPTSNDPVE WESLPNGCPV DSSIHHLLPH ESVCDKYYAC DNGRLVEIGC ASGTHFSPAQ QVCTWPHEAG 280
CEHWTGGGGC STPGDNGGSC GGSTAVPISP TTPVSNISTT PVPDNITTTT TKTPETTTPV SDNSTTPVYE 350
ITTTPVSDNS TTPVSDNSTT PVYEITTTPV SDNSTTPVSD NSTTPVYEIT TTPVSDNSTT PVSDNSTTPV 420
YEITTTPVSD NSTTPVSDNS TTPVYEITTT PVSDNSTTPV SDNSTTPVYE ITTTPVSDIE TTTTTKTPEV 490
STTTTSAPDC DTEGTVTEDP ITEDSTPESN SSEEVPAPAD PELPGTTTNS PPAPSCPECP TVPLTPAEKC 560
KQGCNVAPWA HAECDKYYSC IGNEFRLNIC SEGLHFNPST LTCDFICNAG CDRNIPQVTR HEDGMLIFVP 630
HSFNNKANML ELIEHELNEE F                                                     651
```

```
McIIM4.1  D.....L.....K.....G.....H.....DNSTTPVYEITTTPVSDNSTTPVS.....E.....V.....T.....E
McIIM4.2  D.....L.....E.....E.....L.....DNSTTPVYEITTTPVS.....E.....A.....I.....G
McIIM4.3  G.....P.....K.....G.....L.....GNSTTPVYEITTTPVS.....E.....A.....I.....E
McIIM4.4  D.....L.....K.....G.....L.....DNSTTPVYEITTTPVS.....E.....V.....T.....E
McIIM4.5  D.....L.....K.....G.....L.....D.....A.....I.....E
McIIM4.6  D.....L.....K.....G.....L.....E.....A.....I.....E
```

Figure 5.2 Amino acid sequence of the *Mamestra configurata* insect intestinal mucins (McIIMs). Signal peptides (bold, italics, underlined), chitin binding domains (light blue background) including conserved cysteines (white lettering on dark blue background), additional cysteines (bold, underlined), mucin domains (dark green background) including threonines or serines predicted to be *O*-glycosylated (white lettering) are shown. The alignment below McIIM4 shows the differences between alleles relative to McIIM4.1. The sequences of McIIM2, McIIM3 and McIIM4 have been deposited in GenBank as FJ670567, FJ670568 and FJ670569, respectively.

The McIIMs were rich in threonine (12.4-34.9%), proline (4.3-12.9%), and alanine (6.1-13.6%) with these three amino acids accounting for 53%, 36.1% and 30.2% of the total amino acids in McIIM2, McIIM3 and McIIM4, respectively (Appendix B). Aromatic amino acids accounted for 6.5%, 11.9% and 6.4% of the amino acids in McIIM2, McIIM3 and McIIM4, respectively. As expected, the CBDs were rich in cysteines which ranged from 10.3% to 15.4% (Appendix B). Other amino acids which had a higher than predicted prevalence relative to random occurrence (5%) were asparagine (up to 13%), alanine (9.1%), aspartic acid (14%),

proline (9.3%), histidine (11.1%), leucine (9.3%) and glutamic acid (9.3%). The hybrid CBD-MDs from McIIM3 had a high content of threonine (23.6%), serine (10.9%) and tyrosine (10.3%). The MDs were rich in threonine (up to 68.7%), serine (13.5%) and proline (18.1%). Depending on the sequence of the repeat, MDs could also be rich in valine (up to 9.9%), glutamic acid (17.6%) and alanine (14.3%). A unique characteristic of all MDs was the presence of tandem repeats of varying length and amino acid sequence. For example, "TAAP" was repeated 22 times in McIIM1 MD1, "TTTTTTPAPTT" 8 times in McIIM2 MD1 and "STTPV" 12 times in McIIM4 MD1. McIIM3, with its hybrid MD/CBD had only two "TYTC" repeats as additional repeats would likely have disrupted the integrity of the CBD.

While the carboxy termini of the McIIM were generally hydrophobic, none were predicted to have a glycosylphosphatidylinositol (GPI) membrane anchor sequence or a transmembrane domain. Many of the serines, threonines and tyrosines within the MDs and CBDs were predicted to be phosphorylated; however, PM proteins were not recognized by either an antiphosphothreonine or an antiphosphotyrosine antibody which indicated that this was not the case (data not shown).

5.3.2.3 Gene expression

Northern blot analysis showed that all *McIIM* genes were expressed only in the midgut epithelium (Figure 5.3). The *McIIM2* transcript was approximately 1.1 kb which is in agreement with the 1083 bp cDNA. A smaller, less abundant transcript of approximately 0.8 kb was also observed. The smaller transcript could be an alternatively spliced mRNA, though this was not detected among the cDNAs or 5'RACE products. This may also be the result of cross-hybridization to the *McIIM3* transcript that had regions of similarity to *McIIM2*. The *McIIM3* transcript was approximately 0.8 kb which is in agreement with the 796 bp cDNA. Northern blot analysis using the *McIIM4* cDNA as a probe detected transcripts of approximately 2.8 and 3 kb. Given the similarity between regions of *McIIM4* and *McIIM1*, the smaller transcript could be *McIIM4* (2809-2881 bases), whereas the larger one is likely *McIIM1* (2960 bases) (Shi et al., 2004). Alternatively, the two transcripts may represent differentially spliced forms of the *McIIM4* mRNA. RT-PCR analysis indicated that all *McIIM* genes were expressed in the midgut epithelium of early, middle and late feeding larvae and 24 h starved and molting larvae (Figure 5.3). The levels of *McIIM2* and *McIIM4* transcripts were similar relative to the control gene

McTUB; however, more cycles (40) were required to obtain the similar amount of product from *McIIM3* transcripts indicating that this gene was expressed at a lower level.

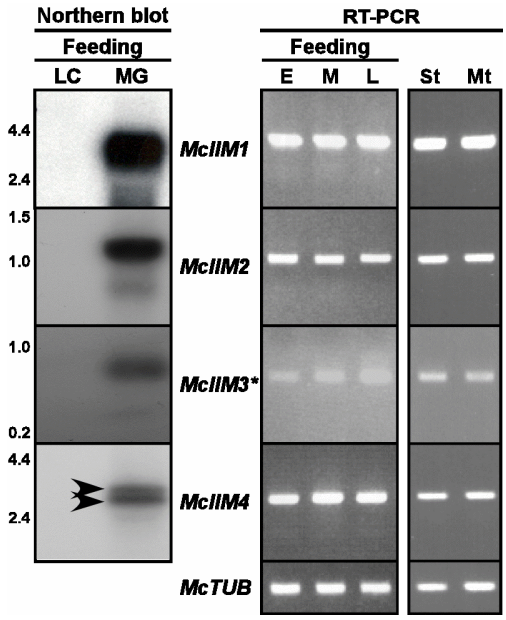


Figure 5.3 *Mamestra configurata* insect intestinal mucin (McIIM) gene expression in feeding, starved and molting larvae. Total RNA from the midgut (MG) or remaining larval carcass (LC) of 5th instar feeding larvae was used in the northern blot analysis. The sizes of the molecular weight markers (kb) are shown in the left hand margin of the northern blot panel. Total RNA from the midgut of early (E), middle (M) and late (L) 4th instar feeding larvae, larvae starved for 24 h (St), and Phase II molting (Mt) (4th to 5th) larvae was used for RT-PCR analysis with 28 or 40(*) cycles. Amplification of the *M. configurata* tubulin (*McTUB*) gene was used as a control in the RT-PCR to ensure equivalent RNA template in samples.

5.3.2.4 Localization

On denaturing SDS-PAGE gels, the recombinant McIIM2 (rMcIIM2) expressed in *E. coli* BL21 (DE3) migrated as a band of ~65-67 kDa (Figure 5.4A) which was unexpected given that the predicted size of the unprocessed form was 31.5 kDa. In 5th instar feeding and 24 h starved larvae or larvae molting from the 5th to 6th instar, the anti-McIIM2 antiserum reacted strongly with a protein of approximately 130 kDa in both PM and midgut (Figure 5.4B). In feeding larvae, two weaker bands with higher molecular weights (≥170 kDa), which may represent the fully glycosylated forms, were also detected in the strongly associated PM fraction (Figure 5.4B). A band of approximately 65 kDa, which may represent the unprocessed McIIM2, was detected in the midgut tissues from all samples, though it was less prominent in the starved and

molting larvae (Figure 5.4B). A lower molecular weight band was also detected in the midgut tissue from feeding larvae, which may be a form in the process of glycosylation or a degradation product (Figure 5.4B).

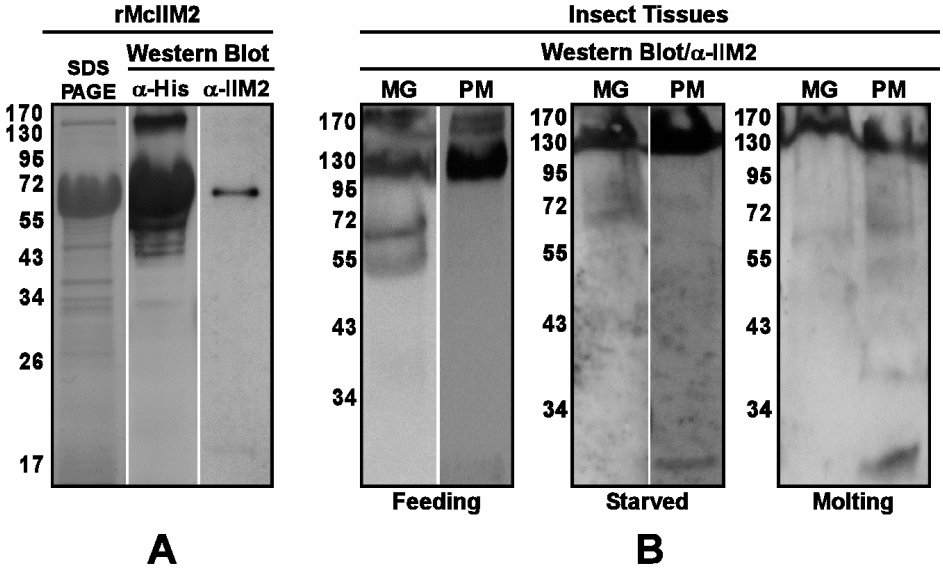


Figure 5.4 Localization of *Mamestra configurata* McIIM2. A:SDS-PAGE and western blot analysis of recombinant *M. configurata* IIM2 (rMcIIM2) expressed in *E. coli*. The blot was probed with antihexahistidine (α-His) antibody and anti-rMcIIM2 (α-IIM2) antiserum. B: Western blot analysis of McIIM2 in the peritrophic matrix (PM) and PM-free midgut (MG) from 5th instar feeding larvae, larvae starved for 24 h and Phase II molting (5th to 6th instar) larvae by western blot analysis. The sizes of the molecular weight markers (kDa) are shown in the left hand margin.

Localization of McIIM3 was not possible as neither of the peptide based antisera reacted with any bands from the PM or midgut extracts (data not shown). Additionally, McIIM3 could not be expressed in any of three *E. coli* strains tested and thus the quality of the antisera could not be evaluated.

The anti-McIIM4 antiserum reacted with a single protein (≥170 kDa) in the strongly associated PM fraction from feeding or starved larvae (Figure 5.5). This may represent the fully glycosylated form as the predicted molecular weight of the unprocessed protein was 65.5-68 kDa. In contrast, this form was not detected in the PM from molting larvae; instead, two proteins of approximately 70 kDa and 78 kDa were detected. The molecular weights of the former was similar to those predicted for the unprocessed proteins from the *McIIM4* alleles (65.5-68 kDa). The McIIM4 profiles from the midguts of all samples were almost the same (Figure 5.5) and

may contain partially processed intermediates or proteins being secreted at the time of dissection. Indeed, the blot from the PM protein samples of feeding larvae revealed additional proteins of 54, 70, 78 and 88 kDa (Figure 5.5), which become more visible the longer the blot is exposed (data not shown). The 54 kDa protein may be an unprocessed form derived from the smaller transcript that was detected by northern blot analysis (Figure 5.3), while the 95 kDa protein may be a partially glycosylated intermediate.

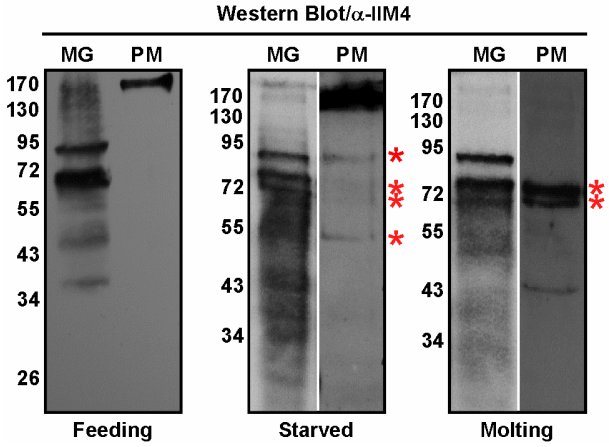


Figure 5.5 Localization of *Mamestra configurata* McIIM4. Protein samples from the peritrophic matrix (PM) and PM-free midgut (MG) of 5th instar feeding larvae, larvae starved for 24 h and Phase II molting (5th to 6th instar) larvae were examined by western blot analysis. Partially processed or nonprocessed forms (red *) are indicated. The sizes of the molecular weight markers (kDa) are shown in the left hand margin.

5.3.3 Chitin Deacetylase 1 (McCDA1)

5.3.3.1 Identification

5' RACE analysis of *McCDA1* revealed a 1263 bp transcript, which was 5 bp longer than the longest cDNA identified in Chapter 4; however, the ORF did not change (390 amino acid). The predicted molecular weight (42.2 kDa) for McCDA1 was in accordance with that observed after 2D PAGE analysis of the loosely associated PM protein fraction (Figure 5.6). LC-MS/MS of this spot revealed six unique peptides (Appendix C) that covered 23% of McCDA1. A 42 kDa band was also detected in the strongly and loosely associated PM protein fraction after 1D separation (Figure 5.6). Additionally, a prominent 31 kDa band resolved into two bands of 32 and 30 kDa when less protein was loaded and the running time for the 1D gel was increased (Figure 5.6). LC-MS/MS peptide sequence data indicated that the 32 kDa band also contained McCDA1, which may be a processed or degraded form (Appendix C).

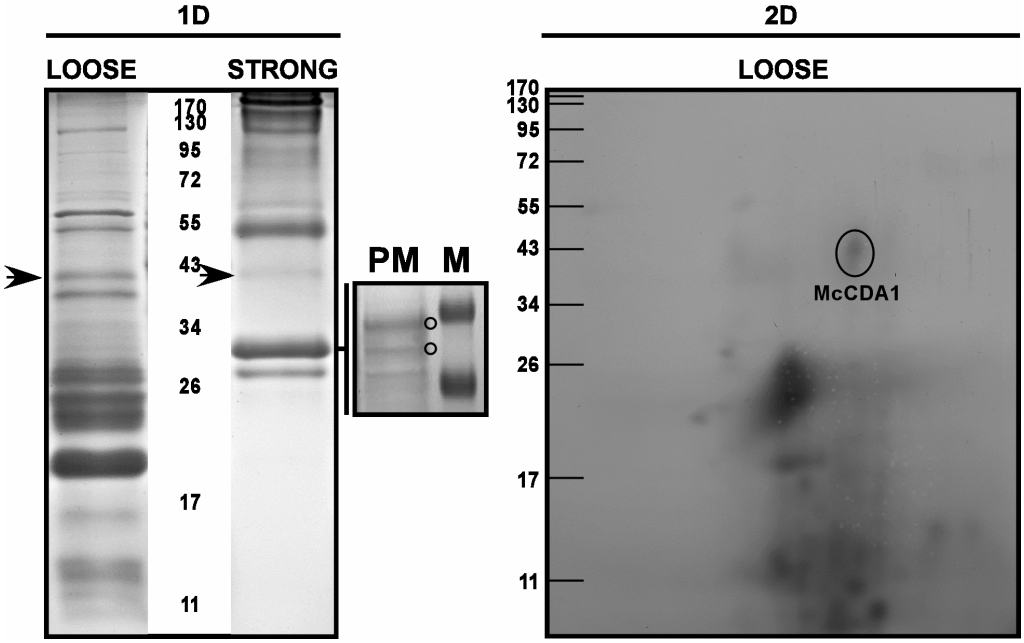
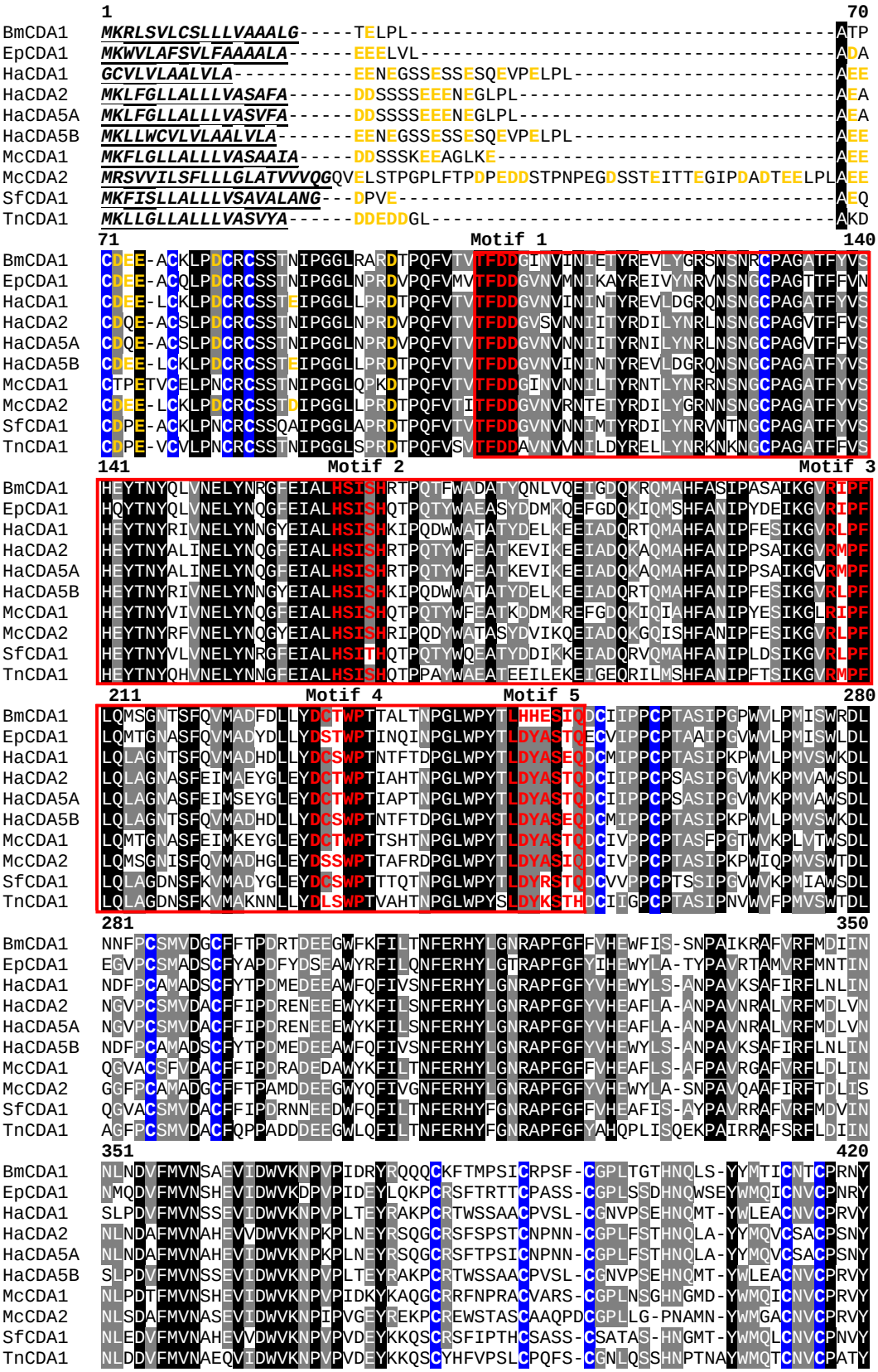


Figure 5.6 One dimensional (1D) and two dimensional (2D) gel electrophoresis of loosely (LOOSE) and/or strongly (STRONG) associated peritrophic matrix (PM) proteins from 5th instar *Mamestra configurata* larvae. The sizes of the molecular weight markers (kDa) are shown between the two lanes in the 1D gel and in the left hand margin in the 2D gel. The protein identified as McCDA1 is indicated with an arrow in the 1D gel and with a circle in the 2D gel. The inset panel to the right of the 1D gel panel shows improved separation of the diffuse 31 kDa band into a 32 kDa and 30 kDa band (circles) when less protein was loaded and the run time for the 1D gel was increased. The 34 kDa and 26 kDa markers are shown (M). Loosely associated proteins were solubilized with Ringer's physiological solution and strongly associated proteins solubilized with 2.5% SDS, 5% β-mercaptoethanol with 500 mM NaCl at 100°C for 5 min.

The amino terminus of McCDA1 and McCDA2 as well as those from other lepidopteran CDAs including, *Bombyx mori* BmCDA1 (assembled from the sequences with GenBank accession numbers, BY917487 and AU002310), *Epiphyas postvittana* EpCDA1 (EV809282), *Helicoverpa armigera* HaCDA1 (EU325545), HaCDA2 (ABU98616), HaCDA5A (GQ411190), HaCDA5B (GQ411191), *Spodoptera frugiperda* SfCDA1 (assembled from Sf2M01668-5-1, Sf2M03887-5-1 and Sf2M03505-3-1), and *T. ni* TnCDA1 (AY966402) were relatively more acidic (e.g., $D^{18, 19}$, $E^{25, 26, 31, 33, 34}$ in McCDA1) than the carboxy terminus (Figure 5.7). McCDA1 had two putative *O*-linked glycosylation sites at T^{211} and T^{227} and one putative *N*-linked glycosylation site at N^{180}. Fifteen cysteines were present in three clusters of five situated between amino acids 35 to 95, 195 to 255 and 344 to 377 (Figure 5.7). The arrangements of the cysteines did not conform to any of the CBD motifs reported for peritrophins.


Motif 1
Motif 2
Motif 3
Motif 4
Motif 5

(Figure 5.7 continued)

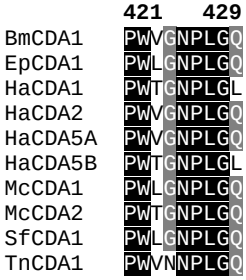


Figure 5.7 Alignment of amino acid sequences of lepidopteran chitin deacetylases (CDAs) that have been demonstrated or proposed to be associated with the peritrophic matrix. The locations of the signal peptide (bold, italics, underlined), acidic amino acids in the amino terminus (yellow lettering), CDA domain (red box) containing the five putative catalytic motifs (red lettering), conserved cysteines (blue background), conserved (gray background) and identical amino acids (black background) are shown. GenBank accession numbers are: *Bombyx mori* BmCDA1 (assembled from BY917487 and AU002310), *Epiphyas postvittana* EpCDA1 (EV809282), *Helicoverpa armigera* HaCDA1 (EU325545), HaCDA2 (ABU98616), HaCDA5A (GQ411190), HaCDA5B (GQ411191), *Mamestra configurata* McCDA1 (EU660852), McCDA2 (HM357864), *Spodoptera frugiperda* SfCDA1 (assembled from Sf2M01668-5-1, Sf2M03887-5-1 and Sf2M03505-3-1), and *Trichoplusia ni* TnCDA1 (AY966402).

5.3.3.2 Phylogenetic analysis and conservation of catalytic motifs in chitin deacetylases

Insect CDA-like proteins were classified into five groups based on phylogenetic analysis and all midgut CDAs clustered in Group 5 (Dixit et al., 2008). However, only seven midgut CDAs (*T. castaneum* TcCDA6-9, *Drosophila melanogaster* DmCDA9, and *Anopheles gambiae* AgCDA9) including one lepidopteran CDA (*T. ni* TnCDA9) were examined in this analysis. Here, I expanded the analysis to include 10 additional CDA proteins from lepidopterans (*M. configurata* McCDA1 and McCDA2, *B. mori* BmCDA1, *E. postvittana* EpCDA1; *H. armigera* HaCDA1, 2, 5A and 5B, *S. frugiperda* SfCDA1) and a dipteran (*D. pseudo-obscura* DpCDA1). The expanded analysis revealed two major clades (Figure 5.8A). Clade I contained two subclades. Subclade Ia contained the lepidopteran midgut CDAs including McCDA1 and McCDA2, SfCDA1, HaCDA1, 2, 5A and 5B, EpCDA1, BmCDA and TnCDA. Subclade Ib contained the *T. castaneum* nonmidgut CDAs, TcCDA1, 2A, 2B, 3, 4, 5A and 5B. Clade II contained the nonlepidopteran midgut CDAs, TcCDA9, DmCDA1 and DpCDA1.

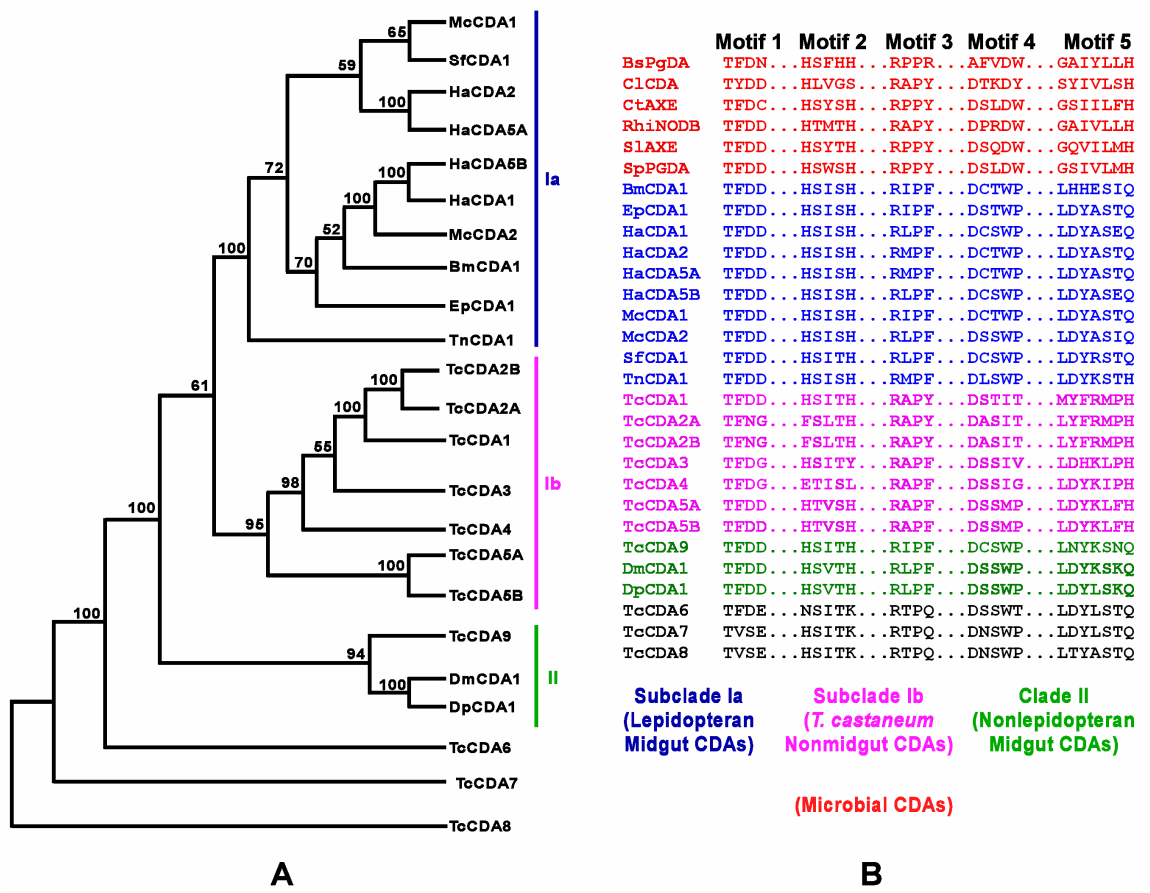


Figure 5.8 Phylogenetic analysis of insect chitin deacetylases (CDAs) (A) and comparison of their putative signature motifs with those of active microbial CDAs (B). Insect CDAs used are from the coleopteran *Tribolium castaneum* TcCDA1 (GenBank accession number, ABU2522), TcCDA2A (ABU25224), TcCDA2B (ABU25225), TcCDA3 (ABW74145), TcCDA4 (ABW74146), TcCDA5A (ABW74147), TcCDA5B (ABW74148), TcCDA6 (ABW74149), TcCDA7 (ABW74150), TcCDA8 (ABW74151) and TcCDA9 (ABW74152); dipterans *Drosophila melanogaster* DmCDA1 (NP_611192), and *D. pseudo-obscura* DpCDA1 (EAL26380), and lepidopterans *Bombyx mori* BmCDA1 (assembled from BY917487 and AU002310), *Epiphyas postvittana* EpCDA1 (EV809282), *Helicoverpa armigera* HaCDA1 (EU325545), HaCDA2 (ABU98616), HaCDA5A (GQ411190), HaCDA5B (GQ411191), *Mamestra configurata* McCDA1 (EU660852), McCDA2 (HM357864), *Spodoptera frugiperda* SfCDA1 (assembled from Sf2M01668-5-1, Sf2M03887-5-1 and Sf2M03505-3-1), and *Trichoplusia ni* TnCDA1 (AY966402). Representative CDAs from microorganisms are *Bacillus subtilis* (BsPgDA, GenBank accession number, BAI84314), *Clostridium thermocellum* (CtAXE, EEU01555), *Colletotrichum lindemuthianum* (ClCDA, 2IW0_A), *Sinorhizobium meliloti* (RhiNODB, ABD67421), *Streptococcus pneumoniae* (SpPGDA, 2C1G_A), *Streptomyces lividans* (SlAXE, ZP_05526521).

CDAs examined in the phylogenetic analysis were also checked for five catalytic motifs identified in the active site of several microbial CDAs from *Streptococcus pneumoniae* (SpPGDA), *Clostridium thermocellum* (CtAXE), *Streptomyces lividans* (SlAXE), *Sinorhizobium*

meliloti (RhiNODB) and *Colletotrichum lindemuthianum* (ClCDA) (Figure 5.8B) (Blair et al., 2005; 2006). Motif 1 is composed of "TFDD" in midgut CDAs from Clade Ia and Clade II and conforms to the consensus sequence for active microbial CDAs. This motif is slightly different in Clade Ib *T. castaneum* CDAs "TF(D/N)(G/D)". Motif 2 is composed of "HSI(S/T)H" in Clade Ia and "HS(I/V)TH in Clade II midgut CDAs and is within the consensus of active microbial CDAs except the isoleucine. This motif is quite variable in Clade Ib nonmidgut CDAs "(H/F/E)(S/T)(I/L/V)(T/S)(H/Y/L)". Motif 3 is composed of "R(I/L/M)PF" in Clades Ia and II midgut CDAs in which the second and fourth amino acids are different than that of the consensus Motif 3 of active microbial CDAs. In contrast, Motif 3 of Clade Ib nonmidgut *T. castaneum* CDAs "RAP(Y/F)" is within the consensus except for the phenyalanine. The consensuses of Motifs 4 and 5 were different than those of Motifs 4 and 5 of active microbial CDAs; however, each motif within the same subclade or clade was similar.

5.3.3.3 Gene expression

Northern blot analysis showed that the *McCDA1* transcript was approximately 1.3 kb which is in agreement with the 1263 bp cDNA (Figure 5.9A). A smaller, less abundant transcript of approximately 800 bp was also observed. Expression was specific to the midgut and RT-PCR detected *McCDA1* transcripts in the midgut from early, middle and late feeding larvae and during molting phases I, II and III (Figure 5.9B), indicating that *McCDA1* is expressed constitutively in the midgut throughout larval development.

5.3.3.4 Localization

Western blot analysis revealed that McCDA1 was present in both the loosely and the strongly associated PM protein fractions; however, the protein was more prevalent in the loosely associated fraction (Figure 5.10). The 42 kDa protein detected by the anti-McCDA1 antiserum was in agreement with the predicted molecular weight of McCDA1. A 42 kDa protein was also detected in the midgut protein fraction, albeit only as a faint band. The anti-McCDA1 antiserum also recognized a 32 kDa protein in the loosely associated PM protein fraction. McCDA1 was also identified by LC-MS/MS peptide sequence analysis from the 32 kDa band (Appendix C) and this protein may be encoded by the smaller, less abundant transcript that hybridizes to the *McCDA1* probe. The anti-McCDA1 antiserum cross-reacted with two proteins of approximately

70 and 72 kDa in the sample from the larval carcass when NaCl was included in the protein extraction buffer. However, the antiserum did not recognize a 42 kDa protein from the foregut, hindgut, Malpighian tubules, tracheae or integument (data not shown), indicating that McCDA1 is associated only with PM and midgut. The presence or absence of 0.5 M NaCl in the extraction buffer did not affect the release of McCDA1 from the PM.

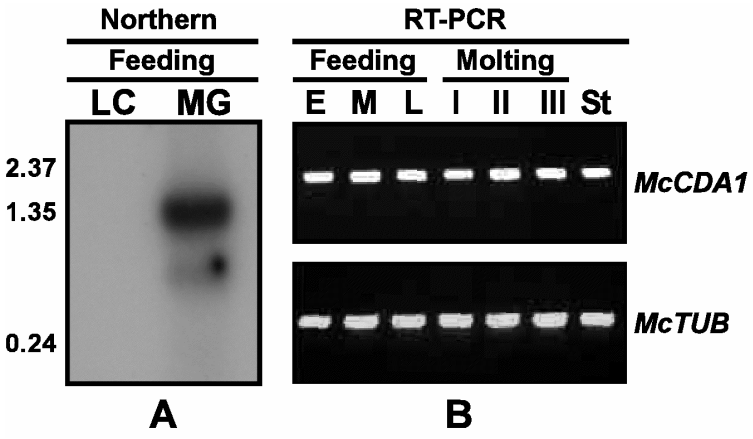


Figure 5.9 Expression of *Mamestra configurata McCDA1*. A: Northern blot analysis of total RNA from the midgut (MG) or remaining carcass (LC) of 5th instar feeding larvae. The sizes of the molecular weight markers (kb) are shown in the left hand margin of the panel. B: RT-PCR analysis of midgut RNA from early (E), middle (M) and late (L) feeding 4th instar larvae, Phase I, II and III molting larvae (from 4th to 5th instar), and larvae starved for 24 h (St). Amplification of the *M. configurata* tubulin (*McTUB*) gene was used as a control in the RT-PCR to ensure equivalent RNA template in all samples.

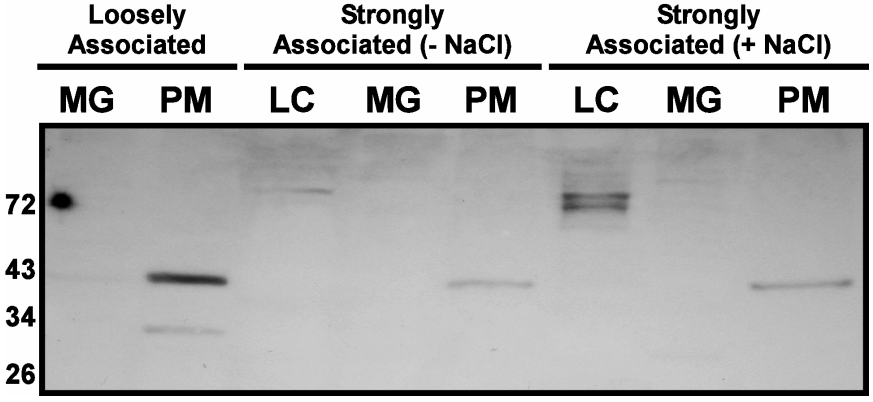


Figure 5.10 Localization of *Mamestra configurata* McCDA1 by western blot analysis. Loosely associated proteins were extracted with Ringer's physiological solution and strongly associated proteins were extracted with 2.5% SDS, 5% β-mercaptoethanol with or without 500 mM NaCl from the peritrophic matrix (PM), midgut (MG) or larval carcass (LC). The sizes of the molecular weight markers (kDa) are shown in the left hand margin.

5.3.3.5 Demonstration of chitin deacetylase activity

Recombinant McCDA1 was expressed in *E. coli* and tested for CDA activity. Three different strains of *E. coli* were tested, BL21, Rosetta 2 and Rosetta gami 2 pLysS. The Rosetta 2 strain expresses tRNA for seven codons that are used rarely in *E. coli*. Rosetta gami 2 pLysS expresses these same tRNAs and has mutations in the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes which should enhance disulphide bond formation; however, this strain grew very slowly. Expression of McCDA1 was low in all strains, but highest in the Rosetta 2 strain as determined by western blot analysis (data not shown). The majority of the protein was present in the inclusion body fraction. After separation by nondenaturing PAGE, the proteins were transferred to a chitin embedded polyacrylamide gel either by passive diffusion as recommended by Trudel and Asselin (1989) or electrophoretically. The gels were stained with FB28 with hyper-fluorescent areas being indicative of conversion of chitin into chitosan. No activity was observed when the proteins were transferred using the passive diffusion process. The electrophoretic method transfered a greater proportion of the proteins to the chitin embedded gel and activity was detected as a diffuse band migrating around 26 kDa on the nondenaturing polyacrylamide gels (Figure 5.11A). A negative control consisting of an *E. coli* extract expressing McPPAD3 did not show CDA activity.

The loosely associated midgut protein fraction extracted with Ringer's solution also showed the same 26 kDa diffuse band as the recombinant McCDA1 (Figure 5.11A) indicating that a similar or the same active chitin deacetylase is also present in the midgut. The loosely associated PM protein fraction, which also contains lumenal proteins (as does the midgut preparation), did not show this band. When the chitin embedded gel was incubated for an extended period (4 days), the loosely associated midgut protein fractions showed two weak additional CDA activities which migrated at about 50 and 70 kDa on the nondenaturing gels (data not shown).

To verify that the 26 kDa band on the nondenaturing gel was McCDA1, several western blot analyses were conducted. Under denaturing conditions, the recombinant McCDA1 migrated as a 42 kDa protein as expected. The antihexahistidine antibody also recognized additional lower molecular weight products in the inclusion body fraction (Figure 5.11B). Under nondenaturing conditions, the antihexahistidine antibody recognized a 38 kDa protein in the recombinant McCDA1 preparations, but also reacted with a protein of 26 kDa and less so with a 17 kDa

protein in the inclusion body fraction. The anti-McCDA1 antiserum recognized a 38 kDa protein in the recombinant McCDA1 preparations, as well as in the loosely associated midgut and PM samples. This antiserum also reacted with a 26 kDa protein in the recombinant McCDA1 inclusion body fraction but not in the midgut or PM samples. The active form of the recombinant McCDA1 appeared to migrate as a 26 kDa protein under nondenaturing conditions; however, it is not clear whether this corresponded to the 26 kDa activity observed in the midgut fraction.

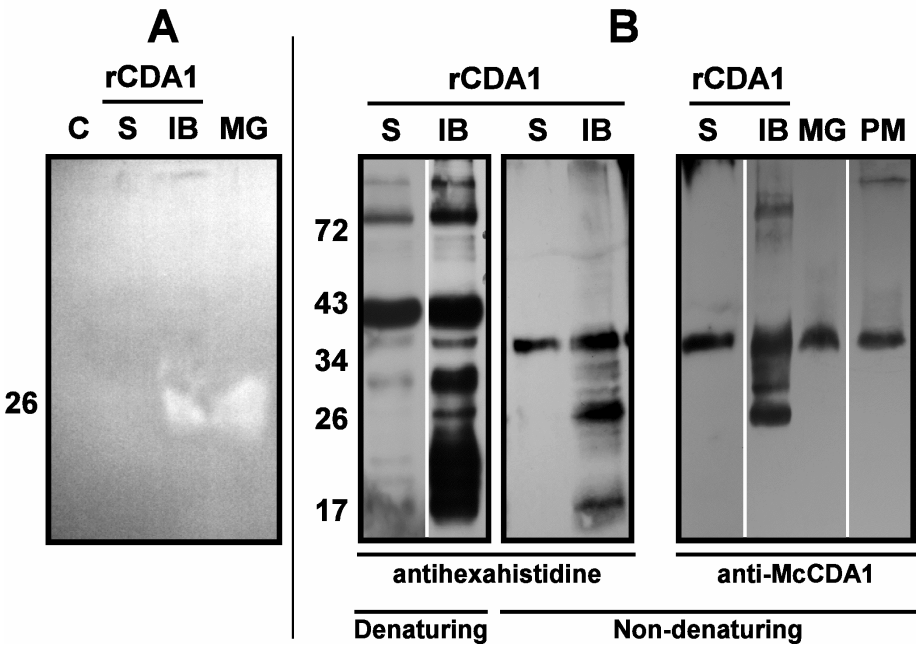


Figure 5.11 Demonstration of chitin deacetylase (CDA) activity. A: CDA activity was examined in the soluble (S) and inclusion body (IB) fractions from *E. coli* Rosetta 2 (DE3) cells expressing recombinant McCDA1. The inclusion body fraction from an *E. coli* line producing an unrelated protein was used as a negative control (C). Insect midgut (MG) tissues homogenized with Ringer's physiological solution were also tested. A fluorescent (white) area is indicative of CDA activity. B: Western blot analysis of soluble or inclusion body fractions of recombinant McCDA1 (rCDA1) with antihexahistidine antibody under denaturing or nondenaturing conditions. The same fractions of the recombinant McCDA1 (rCDA1) or midgut (MG) and peritrophic matrix (PM) tissues were also analysed with anti-McCDA1 antiserum under nondenaturing conditions. Molecular weight markers (kDa) are shown in the left hand margin.

5.3.3.6 Chitosan test

Treatment of the intact or protein stripped *M. configurata* PMs with potassium iodide under acidic conditions produced a reddish-brown colour; however, sulphuric acid treatment by itself also caused a similar colour change. PM-free midguts (negative control) also produced

similar colour changes, indicating that this protocol is not specific enough to determine the chitosan content of the PM. In addition, starch as a second negative control was treated with potassium iodide in the presence or absence of sulphuric acid. Although potassium iodide treatment did not produce a reddish-violet colour in the absence of sulphuric acid; a reddish colour was visible following sulphuric acid treatment, indicating that the colour change could be due to reactions other than chitosan potassium iodide interaction.

5.3.4 Insect Intestinal Lipases (McIIL1-3)

5.3.4.1 Identification

As reported in Chapter 4, proteomic analysis revealed eight putative IILs. Two of these, McIIL1 (EU660853) and McIIL3 (EU660855), and an IIL (McIIL2) that was not identified by the previous LC-MS/MS analysis were detected from seven predominant spots after separation of the strongly associated PM protein fraction by 2D gel electrophoresis (Figure 5.12); these were further characterized in this section. It should be noted that the activity of these enzymes was not demonstrated; thus, they are considered only as putative lipid degrading enzymes.

5' RACE confirmed the transcript sizes of cDNA for McIIL1 and McIIL3 (1127 and 1042, respectively). Upon 2D PAGE separation of strongly associated PM proteins, one protein of 32 kDa and two proteins of 30 kDa corresponded to the new IIL (Marchler-Bauer et al., 2009), McIIL2 (EU660854), a 335 amino acid protein encoded by a 1135 bp cDNA (Figure 5.12) (Appendix C). The predicted molecular weight of the mature protein was 34.5 kDa, which was slightly higher than the molecular weight of the isoforms detected in 2D PAGE separation. In addition, three different 33 kDa proteins corresponded to McIIL1 isoforms, which was in agreement with the predicted molecular weight of the mature McIIL1 (33.4 kDa); however, a single 14 kDa protein corresponded to McIIL3 that was not in agreement with the predicted molecular weight of the mature McIIL3 (30.6 kDa).

The 32 kDa region containing at least two lipases in 2D analysis corresponded to a diffuse 31 kDa band in 1D analysis. This band was one of the predominant bands in the strongly associated PM protein fraction that could be resolved into two bands of 32 and 30 kDa when less protein was loaded and run longer as noted in 5.3.3.1 (Figure 5.6). When these bands were examined by LC/MS-MS, the 32 kDa band contained McIIL1, McIIL2, McIIL3 and the 30 kDa band contained McIIL2 and McIIL3 (Appendix C).

McIIL1 was predicted to have a single *N*-linked glycosylation site at N^{296} and three putative *O*-linked glycosylation sites at T^{322}, S^{325} and S^{329}. McIIL2 had one putative *N*-linked glycosylation site at N^{76} but no *O*-linked glycosylation sites. McIIL3 had one putative *N*-linked glycosylation site at N^{87} and no *O*-linked glycosylation sites.

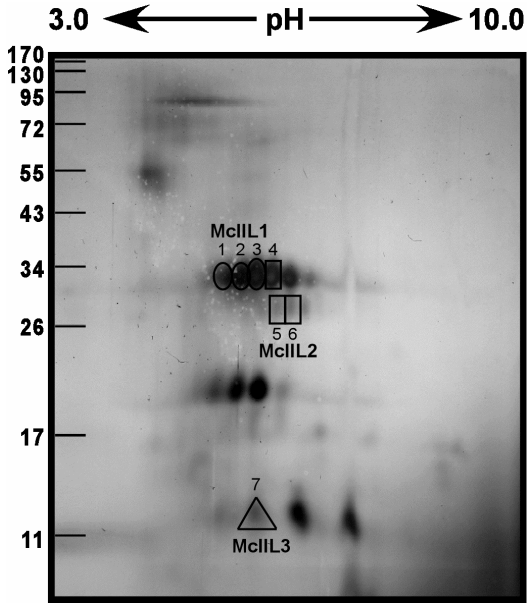


Figure 5.12 Two dimensional gel electrophoresis of strongly associated peritrophic matrix proteins from 5th instar feeding *Mamestra configurata* larvae. The location of proteins identified as McIIL1 isoforms (ovals), McIIL2 isoforms (rectangles), and McIIL3 (triangle) are shown. Numbers of the spots examined (1-7) are shown below or above the spot. Molecular weight markers (kDa) are shown in the left hand margin. Proteins were solubilized with 2.5% SDS, 5% β-mercaptoethanol and 500 mM NaCl at 100°C for 5 min. The pH range in the second dimension is indicated above.

5.3.4.2 Structural features and organization

McIIL1 had two regions corresponding to the conserved lipase active site motif (GXSXG), "GRSVG" (amino acids 156-160) and "GYSVG" (amino acids 181-185). In McIIL2, one region corresponding to a lipase active site was detected, "GHSLG" (amino acids 183-187). Despite having a pancreatic lipase-like domain, McIIL3 did not contain a region corresponding to the lipase active site. As shown in Chapter 4, Figure 4.7, McIILs (except McIIL9) and several other IILs had a conserved arrangement of cysteines, $C^1X_4C^2X_{22}C^3X_{10}C^4$, near the carboxy terminus. This motif includes C^{258}-C^{263}-C^{286}-C^{297} in McIIL1, C^{262}-C^{267}-C^{290}-C^{301} in McIIL2 and C^{226}-C^{231}-C^{254}-C^{265} in McIIL3; however, this cysteine register does not conform to any CBD motifs previously reported for peritrophins.

5.3.4.3 Gene expression

Expression of the three *IIL* genes was specific to the midgut (Figure 5.13A). Transcript sizes for *McIIL1* and *McIIL2* were approximately 1.1 kb, whereas that for *McIIL3* was approximately 1 kb (Figure 5.13A). These sizes were in accordance with the information obtained from 5' RACE and cDNA sequencing. Interestingly, the temporal expression profiles of the individual *M. configurata IIL* genes differed (Figure 5.13B). *McIIL1* was expressed primarily in feeding larvae, while *McIIL3* was expressed primarily during the molt. *McIIL2* was expressed equally both in feeding larvae and during the molt.

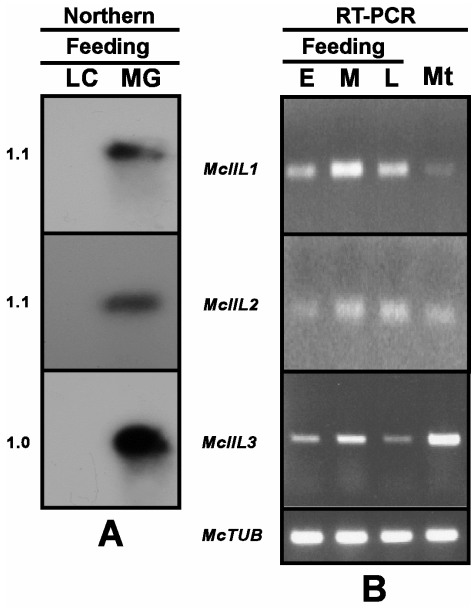


Figure 5.13 Expression of *Mamestra configurata McIIL1*, *McIIL2*, and *McIIL3*. A: Northern blot analysis of total RNA from the midgut (MG) or remaining carcass (LC) of 5th instar feeding larvae. B: RT-PCR analysis of midgut RNA from early (E), middle (M) and late (L) feeding 4th instar larvae, and Phase II molting larvae (from 4th to 5th instar) (Mt). The sizes of the molecular weight markers (kb) are shown in the left margin of Panel A. Amplification of *M. configurata tubulin* (*McTUB*) was used as a control to ensure that equivalent RNA template was present in all samples.

5.3.5 Serine Proteases

5.3.5.1 Identification and localization

The PM protein profiles from feeding larvae contained three prominent bands of approximately 28, 30 and 32 kDa (Figure 5.1), which are within the size range reported for *M. configurata* digestive proteases (Hegedus et al., 2003; Erlandson et al., 2010). Proteomic analysis of the 28 kDa protein band common to the PM of feeding, starved and molting larvae, identified

(peptide mass tolerance and fragment mass tolerance was ±0.3 Da) four trypsins (McSP1, McSP28, McSP30 and McSP31) and five chymotrypsins (McSP13, McSP23, McSP27, McSP35 and McSP41) (Appendix D). One additional trypsin (McSP34) and three additional chymotrypsins (McSP9, McSP29, and McSP38) were identified when the peptide mass tolerance and fragment mass tolerance was set to ±0.6 Da.

5.3.5.2 Gene expression

Two trypsin (*McSP1*, *McSP2*) and two chymotrypsin (*McSP25*, *McSP29*) genes known to be expressed at high levels in feeding larvae were selected for analysis. *McSP1* and *McSP29* encoded proteins that were among those identified in the 28 kDa band (Figure 5.1, Appendix D). All were found to be expressed in the midgut of starved, molting and feeding larvae (Figure 5.14).

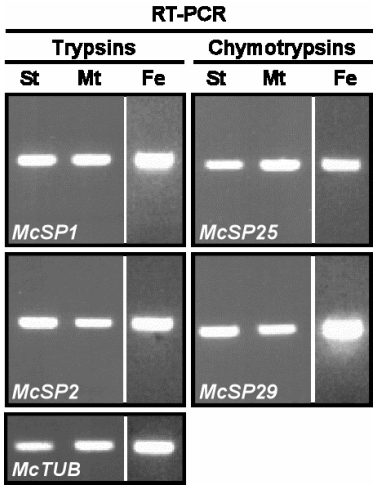


Figure 5.14 Expression of *Mamestra configurata* serine protease genes. Expression of *McSP1* (GenBank accession FJ205402), *McSP2* (FJ205440), *McSP25* (FJ205416), and *McSP29* (FJ205412) in the midgut of 4th instar larvae starved for 24 h (St), Phase II larvae molting from 4th to 5th instar (Mt) and feeding (Fe) larvae is examined by RT-PCR analysis. Amplification of the *M. configurata* tubulin (*McTUB*) gene was used as a control in the RT-PCR to ensure the quality of the RNA template in all samples.

5.3.5.3 Demonstration of serine protease activity in the midgut

Total serine protease activity was examined in the midgut of feeding, starved and molting larvae. Midgut extracts from feeding and molting larvae had similar levels of serine protease activity per midgut, while those of starved larvae had approximately half the level of activity (Figure 5.15). Separation of protease isoforms on nondenaturing SDS-PAGE gels followed by

transfer to a gelatin impregnated polyacrylamide gel showed that the midgut protease profiles were similar for molting, starved and feeding larvae.

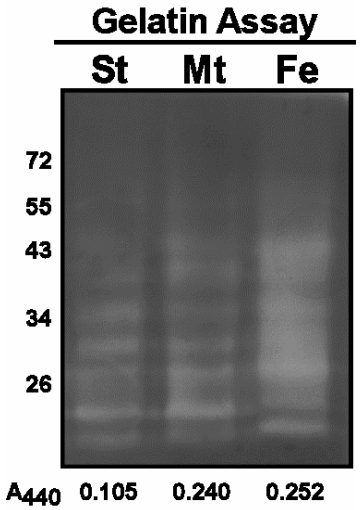


Figure 5.15 Demonstration of total serine protease activity in the *Mamestra configurata* midgut. Serine protease activity was examined in 4th instar larvae starved for 24 h (St), Phase II larvae molting from 4th to 5th instar (Mt) and feeding (Fe) larvae using an in-gel gelatin assay. The total serine protease activity per midgut as measured by the release of dye from an azo-albumin substrate (A_{440} nm) is shown below each lane.

5.4 Discussion

5.4.1 Insect Intestinal Mucins

The defining characteristic of an IIM is the presence of at least one MD containing serines and threonines that are subject to *O*-glycosylation and interspersed with prolines (Wang and Granados, 1997a). As expected, the average threonine, serine and proline contents in the McIIMs were elevated (21.1, 6.2 and 9.9%, respectively). At least one-half of the serines and threonines in McIIM2 and McIIM4 were predicted to be *O*-glycosylated which is comparable to MDs from other lepidopteran IIMs. Less than half of the threonines and none of the serines were predicted to be *O*-glycosylated in McIIM3, though this protein had a unique hybrid CBD-MD and extensive glycosylation would likely impair CBD function. The second hybrid MD in this protein was only 7 amino acids long, but this was within the lower range of that reported for other IIMs; for example, the MDs in peritrophins range from 6 to 1435 amino acids in *Aedes aegypti*, from 14 to 59 amino acids in *T. castaneum*, and from 12 to 108 amino acids in *D. melanogaster* (Venancio et al., 2009). The prolines impart abrupt turns in the polypeptide causing a tightly coiled structure to form with the *O*-linked carbohydrates projecting outward in a

bottlebrush fashion (Van Klinken et al., 1995). This arrangement may prevent digestive proteases from penetrating the PM and shield the underlying CBD. The *O*-linked carbohydrate moieties also impart a mucilaginous property to the protein (Forstner and Forstner, 1994) that may help to lubricate the passage of the food through the digestive tract.

While *O*-linked glycosylation predominates in the IIMs, a small number of *N*-linked glycosylation sites (0, 7 and 12 for McIIM2, McIIM3 and McIIM4, respectively) are also predicted. Recognition of chitobiose (two *N*-acetyl-glucosamine residues) at the core of *N*-linked polysaccharides by a CBD may also contribute to intermolecular interaction between PM proteins as proposed in the model by Shi et al. (2004). Furthermore, *N*-glycans are important for disulphide bond dependent formation of human mucin dimers (Bell et al., 2003), an arrangement that was also proposed for IIMs (Shi et al., 2004).

CBDs comprise a sizeable proportion of the McIIMs and are often flanked or separated by "spacer" regions of 4 to 34 amino acids that are rich in charged and polar amino acids (Appendix E). The spacers separating the internal CBDs of most lepidopteran IIMs have one or two conserved leucines immediately upstream of the CBDs and are replete with charged amino acids and serines. The internal spacers of McIIMs contained a conserved motif (T/V/S/L)L(D/P)NG immediately preceeding the CBD. The *M. configurata* midgut was shown to have CDA activity (Figure 5.11A), suggesting that the PM could also contain deacetylated chitin (chitosan). It is possible that negatively charged groups within the IIM spacers interact with the amino groups of chitosan and help to adsorb IIMs onto the PM, further increasing structural integrity.

The current study has shown that *McIIM* genes are expressed continuously and that the presence of IIMs is a constant feature of the PM despite the fact that overall profiles of the strongly associated PM proteins differ in starved, molting and feeding *M. configurata* larvae. These findings suggest that expression of these genes is not related to feeding status in *M. configurata* larvae. In support of this, Wang and Granados (1998) reported that neonates already have IIM in an amount comparable to that found in feeding larvae. Similarly, expression of *Ae. aegypti AeiMUC1* (Shao et al., 2005), *An. gambiae AgAper14* (Devenport et al., 2005) and *AgAper1* (Devenport et al., 2004) was detected before and after feeding. The continuous expression of IIMs in mosquito guts is intriguing because the PM is formed in response to feeding in mosquitoes. However, Devenport et al. (2004) provided evidence that AgAper1 is

produced and stored in secretory vesicles in midgut epithelial cells until its release into the midgut lumen and incorporation into PM is triggered by a blood meal. As shown in Chapter 3, by midmolt *M. configurata* formed a new, translucent PM covering the inner (former) one. Thus, in the current study, the nonprocessed and partially processed forms of McIIM4 were detected in the PM from midmolt larvae instead of the fully glycosylated form found in the PM of feeding larvae (Figure 5.5), also confirming that the PM at midmolt is a new PM and not a remnant from the feeding larvae. In addition, Tellam et al. (2000) hypothesized that peritrophins bind pre-existing immature PM during the process of PM maturation. The PM in the midmolt may represent immature PM as it contains nonprocessed forms of McIIM4. IIMs may be synthesized even prior to the formation of the immature PM. Indeed, IIM is present within the midgut epithelial cells of early pharate third instar larvae before a PM is formed in *T. ni* (Harper and Granados, 1999).

IIMs are a major component of the PM and likely play a role in protecting it from proteases. As noted above, *O*-glycosylation of serines and threonines within the MDs may physically prevent proteases from reaching the underlying protein. Access to protease cleavage sites may also be restricted by the tertiary structure of the CBD which is held together by three disulphide bridges (Shao et al., 2005). Indeed, TnIIM was rapidly degraded by endogenous digestive proteases after treatment with *O*-glycosidase (Wang and Granados, 1997a) or dithiothreitol, a disulphide bond reducing agent (Wang and Granados, 1997b). Prolines are also common in the spacer regions and may be important in preventing proteolytic attack. As first suggested by Shi et al. (2004), the prolines may allow the spacer elements to form tightly coiled structures that bring the CBDs into close proximity to one another. The glycines are also commonly found in the spacers of some IIMs (e.g., *E. postvittana* IIM1, *H. armigera* IIM1 and *Plutella xylostella* IIM) (Sarauer et al., 2003; Simpson et al., 2007; Campbell et al., 2008); these may facilitate the formation of the tightly coiled structures and effectively limit access of proteases to potential cleavage sites. In addition, the preponderance of negatively charged amino acids in these spacer regions could prevent the interaction of serine proteases with PM proteins.

5.4.2 Chitin deacetylases

CDAs are commonly associated with the PM and phylogenetic analysis revealed that lepidopteran CDAs are closely related to each other. Interestingly, *T. castaneum* CDAs

(TcCDA1-5) encoded by genes expressed in nongut tissues were more closely related to the lepidopteran CDAs that were primarily expressed in midgut than to the TcCDAs expressed primarily in gut (TcCDA6-9). This was in contrast to the finding of Dixit et al. (2008) who reported that all midgut CDAs grouped in a single cluster (Group V); however, their phylogenetic analysis included only one lepidopteran midgut CDA. My analysis is more comprehensive and suggests that midgut CDAs constitute a more divergent group than previously thought. Indeed, the difference in expression patterns and the catalytic motifs between the lepidopteran midgut CDAs and the *T. castaneum* midgut CDAs (TcCDA6-9) support this. For example, *TcCDA6-9* are expressed primarily during larval feeding stages (Arakane et al., 2009), whereas McCDA1 is expressed both in feeding and nonfeeding stages. There are differences in all motifs with the most dramatic differences occurring in Motifs 3 and 5. However, *T. castaneum* nonmidgut CDAs (Clade Ib) also have quite different motifs in relation to lepidopteran midgut CDAs. Although some of the amino acids in the motifs of active microbial CDAs are also present in lepidopteran midgut CDAs (e.g., Motifs 1 and 2), major differences are also apparent especially in Motifs 4 and 5, suggesting that particular functions in relation to catalysis may occur through different homologous amino acids (e.g., replacement of a tyrosine in Motif 3 of microbial CDAs with a phenylalanine in lepidopteran CDAs). The demonstration of CDA activity in *M. configurata* midgut supports this hypothesis.

CDA activity was demonstrated in the *M. configurata* midgut and the recombinant McCDA1 was shown to be active, which is the first demonstration of insect CDA activity. Interestingly, CDA activity was not confirmed for TnPM-P42, an ortholog of McCDA1 (Guo et al., 2005). This could be due to the differences in the protein motifs that contribute to chitin binding and CDA activity or to the differences in the methodologies used in the two studies. For example, TnPM-P42 was expressed in insect cells using a recombinant baculovirus, whereas McCDA1 was expressed in *E. coli* in the current study. Additionally, the CDA activity protocol described by Trudel and Asselin (1990) was modified in my study such that proteins were transferred to the chitin substrate gel electrophoretically rather than relying on diffusion. This greatly improved the efficiency of protein transfer and sensitivity of the assay. By contrast, Guo et al. (2005) followed the CDA activity protocol (Trudel and Asselin, 1990) without modification.

The results of CDA activity assays in *M. configurata* midgut were equivocal. On nondenaturing polyacrylamide gels, the recombinant McCDA1 activity migrated with the main activity associated with the midgut sample at approximately 26 kDa. Both antihexahistidine antibody and anti-McCDA1 antiserum reacted with a protein migrating at 26 kDa when the recombinant McCDA1 was separated under nondenaturing conditions. However, the anti-McCDA1 antiserum did not react with a protein of the same size in the midgut or PM extracts. The anti-McCDA1 antiserum was generated using conjugated peptides from McCDA1 and may not recognize the natural, nondenatured form present in the insect. It is also possible that the activity observed in the insect samples is due to a CDA other than McCDA1.

The demonstration of CDA activity in insect midgut suggests that deacetylated chitin, chitosan, should be present; however, attempts to demonstrate chitosan in PM have so far been unsuccessful. The chitosan detection test that I used, which was described in Peters (1992), is not sufficiently specific. Treatment of PM-free midguts (negative control) and starch (second negative control) also gave a reddish-violet colour in the presence of potassium iodide and sulphuric acid, suggesting the potassium iodide may react with the other polysaccharides or glycoproteins in the PM. Campbell et al. (2008) used a very similar approach based on use of potassium iodide in order to detect chitosan in *H. armigera* PM; however, PM material tested positive for chitosan only after treatment with hot, saturated KOH, indicating that the PM contains a relatively high proportion of chitin, but providing no evidence for endogenous chitosan. Tellam et al. (2000) reported that the alkali-stable PM residue remaining after the extensive protein extraction procedure still contained 28% protein. The remaining proteins or glycans cross-linked to the remaining proteins may mask chitosan or even cross-react with the potassium iodide. More specific methods such as nuclear magnetic resonance spectroscopy may be required to detect chitosan in the PM.

Although genes encoding PM associated CDAs may be expressed in some other tissues, the midgut is the primary site of expression (Guo et al., 2005; Jakubowska et al., 2010) as was demonstrated for *McCDA1* and *McCDA2*. The constitutive expression of McCDAs in the midgut suggests a role in gut physiology throughout development. Deacetylation increases the solubility and decreases the density of chitin fibrils *in vitro* (Cho et al., 2000) and therefore may influence the structure and orientation of chitin fibrils in cuticle (Neville, 1975). Proteins with CDA domains from *D. melanogaster* tracheae were shown to restrict the elongation of tracheal tubes

(Luschnig et al., 2006; Wang et al., 2006). McCDAs may be involved in altering the physical and chemical properties of the chitin in the PM by deacetylation. This may not only alter chitin fibril structure, but may also affect PM integrity and permeability. Supporting this idea, an increase in PM permeability was reported in *H. armigera* larvae infected with a recombinant NPV expressing the midgut HaCDA5a (Jakubowska et al., 2010). Chitin deacetylation may also affect the binding of peritrophins as it is not known whether the CBDs of the peritrophins can interact with chitosan. The continuous expression of *McCDA1* and the upregulation of *M. configurata* chitinase gene during the molt (see Chapter 3) suggest that the putative conversion of chitin into chitosan may be required for the maintenance and protection of the newly formed PM from degradation by the midgut chitinase.

5.4.3 Insect Intestinal Lipases

Lipases have been primarily reported from gut (Simpson et al., 2007; Pauchet et al., 2008), though they may be present in other tissues and organs (Male and Storey, 1981; Smith et al., 1994). While the mechanism by which IILs associate with the PM is unknown, most insect midgut lipases with an esterase-lipase domain, including the McIILs (other than McIIL9), have a conserved register of cysteines ($C^1X_{4\text{-}9}C^2X_{22\text{-}25}C^3X_{10}C^4$) near their carboxy termini. This arrangement of cysteines is part of the esterase-lipase domain, but is found only in insect lipases. The $C^1X_{4\text{-}5}C^2X_{22\text{-}25}C^3X_{10}C^4$ motif is highly conserved among lepidopteran lipases and a variant motif is present in dipteran lipases. The motif was not found in hymenopteran lipases. Conversely, insect midgut acid lipases which have abhydro lipase domain (Position specific scoring matrix Id, 146618) do not have this unique arrangement of cysteines. As well, insect lipases from other tissues such as accessory glands and yolk, or mammalian pancreatic lipases do not have this cysteine motif. The cysteine motif does not conform to the CBDs of peritrophins (Tellam et al., 1999) or cuticle proteins (Rebers and Riddiford, 1988) and may constitute a new type of CBD that would allow association of the insect intestinal lipases with the PM.

The lipase active site comprised a serine-histidine-aspartic acid catalytic triad, in which the serine is a part of a conserved motif (GXSXG) (Brady et al., 1990). McIIL1 has two GYSVG motifs, whereas McIIL2 has a single GHSLG motif. *D. melanogaster* and *E. postvittana* have midgut acid lipases that also contain a “GHSQG” motif (Pistillo et al., 1998; Simpson et al., 2007). In contrast, McIIL3 does not have a comparable motif, suggesting that it may not be

catalytically active. Of the 10 lipase genes reported from *E. postvittana*, at least four of them encode enzymes that have either a glycine or glutamate instead of serine in the active site and are unlikely to be active (Simpson et al., 2007). Christeller et al. (2010) suggested two possible funtions for the inactive lipases. First, they may bind to the hydrolysis products of fats and de-esterified sterols in the lumen and act as carriers to the midgut epithelium for absorption. Second, they may bind to lipase inhibitors present in the diet and prevent active lipases from being inhibited. This is somewhat analogous to the presence of inactive serine proteases in the lepidopteran midgut (Mazumdar-Leighton et al., 2000) where binding of inactive serine proteases by inhibitors was proposed to control the level of random unwanted digestion (Holt et al., 2003).

McIIL1-3 have one putative *N*-linked glycosylation site. *Drosophila* Lip1 has five potential *N*-linked glycosylation sites, whereas *Drosophila* Lip3 has one potential site (Pistillo et al., 1998). In human lipoprotein lipases, *N*-linked glycosylation is required for activity (Semenkovich et al., 1990) and may also be important for the activity of IILs. McIIL1 also has three potential *O*-linked glycosylation sites; however, the importance of *O*-linked glycosylation is not known.

The *McIIL* genes are expressed exclusively in the midgut. The association of McIILs with the PM and the expression of their genes in the midgut suggest they have a role in lipid digestion. Thus, *McIIL1* expression declined during the molt like that of a lipase gene from *B. mori*, *Bmlipase-1* (Ponnuvel et al., 2003). The authors suggested that Bmlipase-1 was not required at the molt because larvae do not ingest food during this stage. In contrast, *McIIL2* was expressed in all developmental stages and *McIIL3* was expressed only during the molt indicating that certain McIILs may have roles other than in digestion as discussed above.

5.4.4 Serine Proteases

Serine protease gene expression and protease activity were detected in the *M. configurata* midgut throughout a larval stadium. In addition, serine proteases were integral components of the *M. configurata* PM in feeding and nonfeeding stages. The presence of serine proteases in the PM may be related to their being trapped by the PM during transit of these small enzymes from the midgut epithelium to the bolus. The presence of serine proteases in the PM could improve the

efficiency of digestion by further hydrolyzing polypeptides as they move through the PM toward the epithelium.

Serine protease activity in the midgut of feeding *M. configurata* larvae was approximately double that in starved larvae. This agrees with the general observation of stimulation of trypsin activity upon feeding by insects. For example, trypsin activity was found to be induced by feeding in adult simuliids (Yang and Davies, 1968). In female mosquitoes, the PM is formed in response to feeding, which corresponds with the release of digestive enzymes (Jiang et al., 1997). The PM may also play a role in nonfeeding larvae, especially during periods of extended food deprivation when serine protease genes are still expressed. In this situation, the PM may serve to maintain lumenal enzyme stocks through immobilization which would allow rapid digestion and assimilation of nutrients once feeding resumes. Although the PM from molting larvae is unique in its physical architecture (as shown in Chapter 3) and biochemical composition (as shown in Chapter 5), it had a 28 kDa protein band that in feeding larvae contained many serine proteases. The reason for high levels of serine protease gene expression and activity in molting larvae is less apparent. As in starved larvae, these enzymes may be produced in advance of the imminent resumption of feeding once the molt is complete.

In conclusion, this study indicated that IIMs, CDAs, and serine proteases are present in the PM and their genes are expressed throughout *M. configurata* larval development. Recombinant McCDA1 is active and CDA activity is present in the midgut. Several putative lipases are also associated with the PM; however, they show differences in their developmental expression patterns.

The studies in Chapter 4 revealed that the PAD was common to all *M. configurata* peritrophins. The following chapter expands this finding and specifically focuses on genes encoding proteins with PADs (PPADs) that are expressed in the midgut. Structural organization of PPADs (including peritrophins) together with the phylogenetic analysis of their PADs is used to examine the evolutionary development of these proteins and classify them in lepidopterans. Finally, modeling of different PAD isoforms will be presented.

6. SURVEY, CHARACTERIZATION AND EVOLUTION OF PROTEINS WITH PERITROPHIN-A DOMAINS[4]

6.1 Introduction

Among the peritrophin chitin binding domains (CBDs), the Peritrophin-A domain (PAD) is by far the most common type, with Peritrophin-B domains (PBDs) and Peritrophin-C domains (PCDs) being found in only a few dipteran peritrophins. PADs have been identified primarily in peritrophins; however, they have also been reported in midgut chitinases (Bolognesi et al., 2005) and integument proteins (He et al., 2007). Cat flea, *Ctenocephalides felis*, adults also have proteins with PADs, PPADs, in their Malpighian tubules, hindgut and rectum (Gaines et al., 2003), although they lack a peritrophic matrix (PM) (Peters, 1992). The genomes of *Drosophila melanogaster* and *Tribolium castaneum* each encode 42 PPADs, showing that in insects this protein family is quite large (Gaines et al., 2003; Jasrapuria et al., 2010).

Though genes encoding PPADs have been shown to be expressed in multiple tissues; most studies have been limited to investigating expression of these genes in midgut or cardia, or localization of peritrophins to the PM. Indeed, lack of information on the distribution of this family of proteins is a gap in our understanding of these proteins. The wide distribution of PPADs across tissues and species raises the question "How are these proteins related to each other and how have they evolved?". However, to date only Shen and Jacobs-Lorena (1999) examined the evolutionary relationship of PADs. Their phylogenetic analysis revealed that PADs in insect chitinases and peritrophins evolved from a common ancestor. A deficiency of their analysis is that it included only five peritrophins: *Anopheles gambiae* adult peritrophin 1 (Shen and Jacobs-Lorena, 1998), *D. melanogaster* peritrophin A (GenBank accession number, AF030895), *Lucilia cuprina* peritrophin-44 (Elvin et al., 1996) and peritrophin-95 (Casu et al., 1997), and *Trichoplusia ni* IIM22 (Wang and Granados, 1997b). In addition to the absence of a comprehensive phylogenetic investigation on PPADs, the three-dimensional structure of the PAD has not been resolved, though it was inferred by comparison to that of tachycitin (Shen and Jacobs-Lorena, 1999; Sarauer et al., 2003; Behr and Hoch, 2005), an antimicrobial chitin binding protein from the horseshoe crab *Tachypleus tridentatus* (Kawabata et al., 1996; Suetake et al., 2000). Tachycitin CBD is a good reference for modeling of PADs because the 6 cysteines present in the PAD align with cysteines found in the tachycitin CBD which contains 10

[4] Part of this chapter is published in *Journal of Insect Physiology* **56:**1711-1720, 2010.

cysteines. In addition, several aromatic amino acids in tachycitin and PADs also overlap and this is important for prediction of the aromatic amino acids involved in ligand binding. Previous characterization studies revealed that lepidopteran peritrophins contain two PAD isoforms, one having a typical six cysteine register and a second type having two additional conserved cysteines within the typical register (Sarauer et al., 2003; Shi et al., 2004). The phylogenetic and structural relatedness of two PAD isoforms is also not known although both types are commonly found in lepidopterans.

In the current study, three proteins with only PADs (McPPAD1-3) are characterized from *M. configurata*. The tissue distribution of McPPAD1-3 is investigated at the gene and protein level. Additionally, PADs from all known lepidopteran peritrophins, as well as those identified in *M. configurata* and *Bombyx mori* midgut EST databases and therefore likely to be associated with the PM, are used in a phylogenetic analysis. This information, in conjunction with the domain organization of the PPADs is used for a classification of these proteins including peritrophins and an evolutionary model is developed. As a final step, protein fold recognition analysis using tachycitin CBD as the reference is used to construct models for the two PAD isoforms associated with lepidopteran peritrophins.

6.2 Material and Methods

6.2.1 Identification of Genes Encoding Proteins Containing Peritrophin Domains in Lepidopterans

Six *B. mori* midgut EST databases (11,948 ESTs in total) were accessed from the NCBI (http://www.ncbi.nlm.nih.gov/UniGene/lbrowse2.cgi?TAXID=7091); Lib. 15160 “swb” (7,252 ESTs), Lib. 19475 “F1mg” (2,611 ESTs), Lib. 19479 “FJssb” (1,261 ESTs), Lib. 19480 “JFsb” (567 ESTs), Lib. 19487 “mg” (257 ESTs), and Lib. 19560 “*B. mori* midgut cDNA library” (19 ESTs). These and the *M. configurata* contigs and singleton sequences were translated for all six reading frames. A Perl script was developed to search for proteins encoded by the mRNAs that contained a PAD with the consensus $CX_{13\text{-}20}CX_{5\text{-}6}CX_{9\text{-}19}CX_{10\text{-}14}CX_{4\text{-}14}C$, a PBD with the consensus ($CX_{12\text{-}13}CX_{20\text{-}21}CX_{10}CX_{12}CX_{2}CX_{8}CX_{7\text{-}12}C$) or a partial PCD with the consensus ($CX_{8\text{-}9}CX_{17\text{-}21}CX_{10\text{-}11}CX_{12\text{-}13}CX_{11}C$) where C is a cysteine and X is any amino acid. All databases were searched for genes encoding proteins with PADs, but only the *M. configurata* databases were searched for genes encoding proteins with PADs, PBDs or PCDs. To be declared a putative

PAD, conserved aromatic amino acids also had to be present between C2 and C3 and/or C4 and C5. ESTs found to encode a putative PAD were aligned using Sequencher (85% identity threshold) and the longest sequence or contig identified. These were compared to the *B. mori* protein database GLEAN in SilkDB (http://silkworm.genomics.org.cn) using BLASTX. The EST was retained if a protein was identified with the same reading frame; otherwise the sequence was inspected manually.

Additional PPADs were obtained by a BLAST search of the NCBI database using the *M. configurata* peritrophins and from publications. The final set of PPADs used for phylogenetic analysis (Appendix F) included *M. configurata* McIIM1 (AY057052), McIIM2 (FJ670567), McIIM3 (FJ670568), McIIM4 (FJ670569), McPM1 (AY277403), McPPAD1 (GU596430), McPPAD2 (GU596431), McPPAD3 (GU596432), McPPAD4 (GU596433); *Plutella xylostella* PxIIM (AF545582); *T. ni* TnIIM22 (AF000606), TnCBP1 (AAR06265); *Helicoverpa armigera* HaIIM86 (ABW98670), HaIIM2 (EU325543), HaIIM3 (EU325564), HaIIM4 (EU325565), HaIIM5 (EU449965); *Epiphyas postvittana* EpIIM1 (EV809288), EpIIM2 (EV809261), EpIIM3 (EV809248), EpPTN1 (EV809224), EpPTN2 (EV806871); *Spodoptera frugiperda* SfPer-33k (contig assembled from AAS89976 + contig 32 provided by Dr. W.R. Terra + Sf2M02530-5-1 from SpodoBase, Nègre et al., 2006), Sf27 (provided by Dr. W.R. Terra); *B. mori* BmIIM1 (BGIBMGA000185-PA), BmIIM2 (BGIBMGA001480-PA), BmPPAD1 (BGIBMGA009641-PA), BmPPAD3 (BGIBMGA003270-PA), BmPPAD4 (BGIBMGA011851-PA); *S. exigua* SeCBP66 (ABW98673), SeIIM (EU047712) and *Loxostege sticticalis* LsCBP (ACQ65651).

6.2.2 Rapid Amplification of cDNA Ends (RACE)

The entire coding region of McPPAD2 was determined using the 5' RACE System (Invitrogen, Carlsbad, CA, USA). The gene specific primer 1 (5'-AAACTGCCGAGGGTGTAGC-3') was used to convert the target mRNA into cDNA and the gene specific primer 2 (5'-AGTAGGGGAGAGTAGGGTC-3') was used to amplify the 5' ends from the cDNA.

6.2.3 Analysis of Gene Expression

Expression of target genes was examined in three stages of feeding (early, mid, and late feeding) and three stages of molting (Phase I, Phase II, and Phase III) using reverse transcription-

PCR (RT-PCR) or northern blot analysis as described in Chapter 2. Primers used in RT-PCR analysis are shown in Table 6.1.

Table 6.1 Primers used in expression analysis of McPPAD genes by RT-PCR.

Primer	Sequence
McPPAD1 Fp	CCACCGTCCTGCTCCTGAAC
McPPAD1 Rp	CAGGTCTGCAGTTGGAAGTC
McPPAD2 Fp	CTTGCAGTGTGGTCGGCAGCT
McPPAD2 Rp	CGTATGATTTAGACAACGGCAG
McPPAD3 Fp	AGCGACTTTGTGACTGACGATATG
McPPAD3 Rp	GCATTGAATCAATCACTAGACATCTCC
McTUB1 Fp	AGCGTACCATCCAGTTCGTG
McTUB1 Rp	TGGCGTACATGAGGTCGAAC

[1]**Abbreviations:** Fp, forward primer; Mc, *M. configurata*; PPAD, proteins with peritrophin-A domain; Rp, reverse primer; TUB, tubulin.

6.2.4 Expression of Recombinant Proteins in Escherichia coli

The protein coding sequence of McPPAD1-3 without the signal peptide was amplified using the One-Step RT-PCR kit (Invitrogen) as described in Chapter 2. The sequences of the primers used in RT-PCR are shown in Table 6.2.

Table 6.2 Primers used to amplify open reading frames of *M. configurata* McPPADs to generate constructs for expression of recombinant proteins in *E. coli.*

Primer	Sequence
McPPAD1 Fp	GGAATTCCATATGGAGATTCCTCAGAAGAACGCCACC
McPPAD1 Rp	CCGCTCGAGTTATTGCTCGTATTTTATGAATTCACT
McPPAD2 Fp	GGAATTCCATATGCAACCCGGTGCAGGTATTGGCGCC
McPPAD2 Rp	CCGCTCGAGTTAGACAACGGCAGCGGTGCGGGGAGC
McPPAD3 Fp	GGAATTCCATATGCGAGTCAGAAAGATGGAATCCTTC
McPPAD3 Rp	CCGCTCGAGCTAGACATCTCCCTTAAACCTAAACAT

Abbreviations: Fp, forward primer; Mc, *M. configurata*; PPAD, proteins with peritrophin-A domain; Rp, reverse primer.

6.2.5 Antisera Production, Protein Extraction and Western Blot Analysis

Antiserum for McPPAD1 was generated using a recombinant protein (rMcPPAD1) and antiserum for McPPAD2 was generated using synthetic peptides (Table 6.3). Antisera for McPPAD3 were generated using both conjugated synthetic peptides (Table 6.3) and recombinant protein (rMcPPAD3).

Table 6.3 Sequences of the conjugated peptides used to generate antisera specific for *M. configurata* McPPAD proteins.

Peptide name[1]	**Company**	**Sequence**
McPPAD2 P1	EzBioLab	GIGAWGGGGGGWGGAGGWGGC
McPPAD2 P2	EzBioLab	CKVGGLQRACTDPTLPYCNAG
McPPAD3 P1	EzBioLab	RNMRNDRQMRRGMFRFKGDV
McPPAD3 P2	EzBioLab	VDPEIEDMEQEQKNEEEMDR

[1] **Abbreviations:** Mc, *M. configurata*; PPAD, proteins with peritrophin-A domain; P1, peptide 1; P2, peptide 2.

PM, foregut, midgut (PM-free), hindgut, Malpighian tubules, fat body, tracheae and integument were dissected from 20 to 60 (depending on the specific tissue) feeding 5th instar larvae and washed in ice-cold protease inhibitor cocktail (Roche Diagnostics GmbH, Indianapolis, IN, USA). Strongly associated proteins were extracted from these tissues as described in Chapter 2. For the western blot analysis of McPPAD1, basal lamina and midgut epithelium were also used. To collect these tissues, 5th instar larvae were flooded with sterile artificial hemolymph solution (SAHS) (Palli and Locke, 1987), and integument, large tracheae and Malpighian tubules were carefully removed. The gut was ligated anterior to the foregut-midgut junction and posterior to the midgut-hindgut junction with silk surgical thread. The midgut was removed by cutting distal to the ligatures. It was transferred to a glass slide with a drop of dispase solution which separates the midgut epithelium from basal lamina and then incubated for 15 min. After removal of the dispase solution, the sample was incubated in SAHS for 30 min. The basal lamina was removed and placed in a drop of ice cold Ringer's solution containing protease inhibitor cocktail. The midgut epithelium was cut lengthwise and the PM and food bolus removed. The midgut epithelium was transferred to a drop of ice cold Ringer's solution containing protease inhibitor cocktail. Both tissues were then transferred to separate 1.5

mL microcentrifuge tubes in liquid nitrogen and stored at -80ºC until protein extraction as described above.

6.2.6 Computational Analysis

One hundred and fifty PADs from the PPADs above (Appendix G) were subjected to phylogenetic analysis using PHYLIP version 3.5C (distributed by J. Felsenstein at www.cbr.nrc.ca/cgi-bin/WebPhylip/index.html). Dendrograms were constructed according to the neighbor-joining method and confidence values for the branches were determined using bootstrap analysis where 100 trees were developed from randomly resampled data generated by CLUSTALW.16 as provided in MUSCLE (www.ebi.ac.uk/Tools/muscle/) (Edgar, 2004)**.** The CLUSTALW alignment was performed using the Blosum scoring matrix with the following gap penalties: opening gap 10, end gap 10, extending gap 0.05 and separation gap 0.05. Initially, the majority rule (branches with bootstrap values >50%) was used to construct a consensus tree for all PADs examined. Subsequently, the extended majority rule (branches with bootstrap values < 50% but which agree with the majority rule) was used to construct a more detailed tree to tease out relationships between the PADs within the major clades. The structure of tachycitin (Protein Data Bank code 1DQC) was used as a template to model PAD structure using Modeller 9 version 6 (http://cbsuapps.tc.cornell.edu/loopp.aspx) with default settings.

6.3 Results

6.3.1 *Characterization of Proteins with only PADs in* M. configurata

6.3.1.1 Identification

Three cDNAs from the *M. configurata* midgut EST database were found to encode proteins containing one or two PADs, one of which (McPPAD1, GU596430) was identified in Chapter 4. The new putative proteins were denoted McPPAD2 and McPPAD3. The *McPPAD3* cDNA contained a complete open reading frame, while additional sequence information was obtained for *McPPAD2* using 5' RACE. A 774 bp cDNA encoded McPPAD2, a 147 amino acid protein with a 21 amino acid signal peptide (Figure 6.1) and a predicted molecular weight of 12 kDa without the signal peptide. A 1285 bp cDNA encoded McPPAD3, a 258 amino acid protein with a 21 amino acid signal peptide (Figure 6.1) and a predicted molecular weight of 27.6 kDa without the signal peptide.

McPPAD1

MIAKFLTTVL LLNVVLTAEI PQKNATETAK FRVFENVDPN DLSCDPAGHI FLLLPHFTDC SKFYMCAHGE **70**
EVEFQCPGGL IFDFQLQTCN WAWDTTCQLR TPQDEDEGSG DEADSLIGIF TDELEHQPVD TVASVRPISP **140**
MQGRYNGIIN CARADAAARQ VPYKGDCQRY WKCVAGVPQV AFCSDGLFFN EHTQQCDFEA NSKCVLSQED **210**
ELQSEFIKYE Q **221**

McPPAD2

MSAKIVLVLC ALIAWVSSVP AQPGAGIGAW GGGGGGWGGA GGWGGAGGWG GAGAGAWGGA GAGVNPWIPP **70**
WVSSAPWYPQ WVPCSVVGSS CVDCNTKVVC TKVGGLQRAC TDPTLPYCNA GECSATPSAV CGGPAVDAVA **140**
PRTAAVV **147**

McPPAD3

MASVRARHLL ILLLVSCVKT DRVRKMESFN SAVFNSLREA SNQTQADTLN LPANASSIRE NITDTFSCEN **70**
RTYGYYADVD NDCQVFHVCL PSQTPSGRNI TYRWSFICPN ETVFNQEVLT CTRVADSIDC KDSPDFYHVN **140**
LEIGKVSNKT EEVKDKKDMR KTQKRKQEKR NQNIVMGSEV SDFVTDDMQA HLEPVELIKR QENGPVLIEA **210**
VIESVDPEIE DMEQEQKNEE EMDRIVMERN MRNDRQMRRG MFRFKGDV **258**

Figure 6.1 Amino acid sequence of the *Mamestra configurata* proteins with peritrophin A domains (McPPADs). Signal peptides (bold, italics, underlined) and peritrophin A domains (light blue background) including conserved cysteines (dark blue background), and additional cysteines (underlined) are shown. The sequences of McPPAD1, McPPAD2 and McPPAD3 have been deposited in GenBank as GU596430, GU596431, GU596432, respectively.

6.3.1.2 Structural features and organization

McPPAD1 had two PADs, while McPPAD2 and McPPAD3 each had one PAD (Figure 6.1). The arrangement of the cysteines comprising the PADs was $C^1X_{15}C^2X_5C^3X_9C^4X_{12}C^5X_7C^6$, $C^1X_9C^2X_5C^3X_9C^4X_{12}C^5X_7C^6$, and $C^1X_{14}C^2X_5C^3X_{18}C^4X_{12}C^5X_8C^6$ for McPPAD1, McPPAD2 and McPPAD3, respectively. The PAD in McPPAD2 had two additional cysteines located six amino acids downstream of C1 and seven amino acids downstream of C5 within the PAD cysteine register.

The McPPADs were predicted to be glycosylated to varying degrees. McPPAD1 had one *N*-linked glycosylation site at N^{24}, but no predicted *O*-linked glycosylation sites. McPPAD2 had two putative *O*-linked glycosylation sites at $T^{126, 143}$, but no *N*-linked glycosylation sites. McPPAD3 had seven putative *N*-linked glycosylation sites at $N^{42, 54, 61, 70, 99, 110, 148}$, but had no *O*-linked glycosylation sites.

CBDs in McPPADs were rich in charged amino acids (Appendix H). However, the regions outside the PADs in McPPAD1 and McPPAD3 also contained a disproportionate number of charged (33% for McPPAD1 and 39% for McPPAD3) and hydrophilic amino acids (Figure 6.2). Conversely, the region on the amino terminal side of the PAD of McPPAD2 contained a high proportion of neutral amino acids (glycine, alanine and valine), while the carboxy terminal region was hydrophobic.

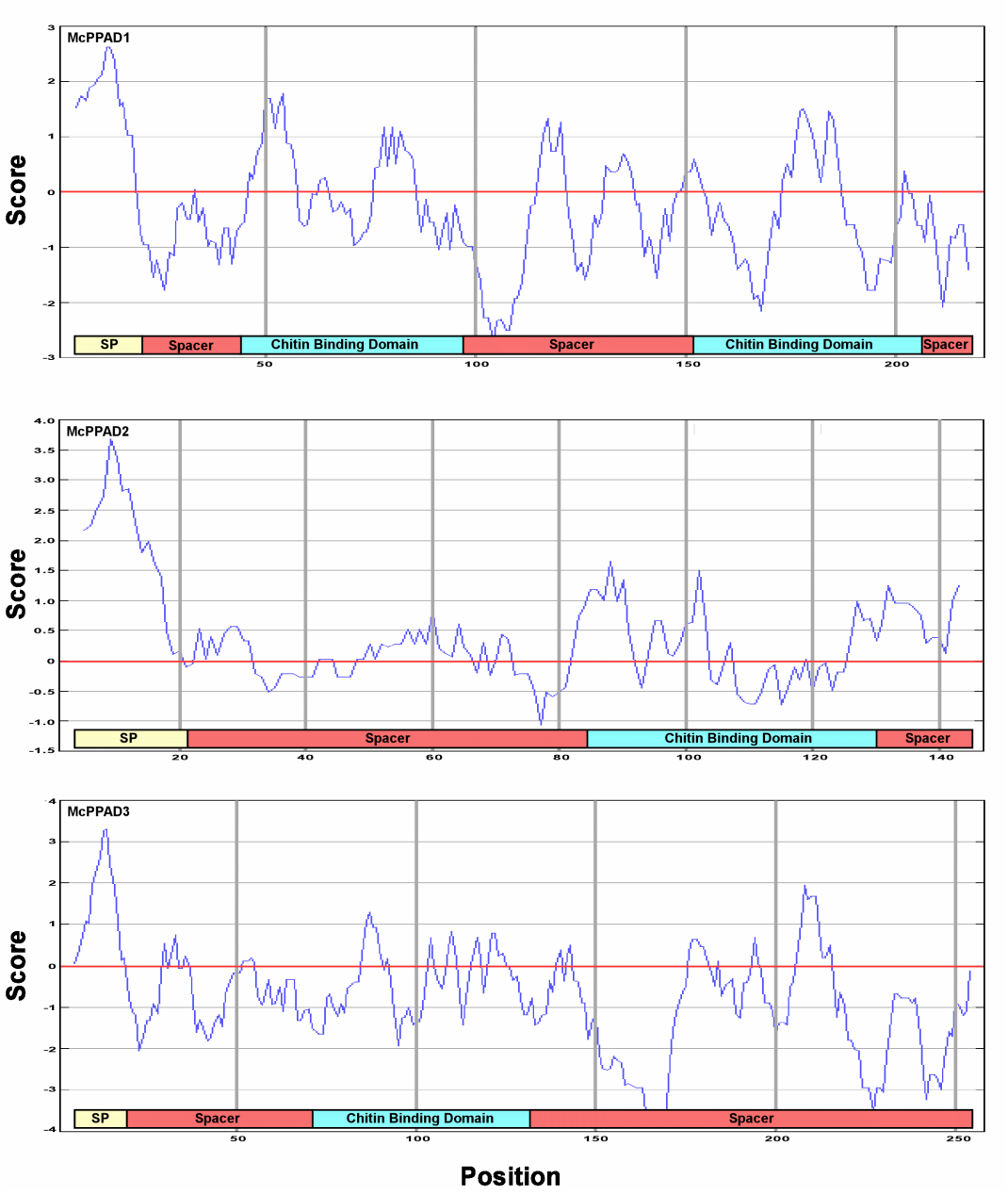


Figure 6.2 Hydrophobicity plots of *Mamestra configurata* proteins with peritrophin A domains (McPPADs). Signal peptides (SP), chitin binding domains and spacer regions are shown as yellow, blue and red boxes, respectively. Hydrophobicity score is according to Kyte and Doolittle (1982) where scores <0 are increasingly hydrophilic and >0 are increasingly hydrophobic.

6.3.1.3 Gene expression

Northern blot analysis using total midgut RNA probed with the *McPPAD1* cDNA detected transcripts of approximately 1.0 and 0.8 kb (Figure 6.3). The 1.0 kb transcript agreed

with the 951 bp cDNA. The *McPPAD2* probe detected four transcripts of approximately 3.0, 0.8, 0.5, and 0.1 kb (Figure 6.3). The major transcript (0.8 kb) corresponded to the 774 bp cDNA. The *McPPAD3* cDNA probe detected a single transcript of approximately 1.3 kb (Figure 6.3) which was in agreement with the 1285 bp cDNA.

RT-PCR analysis indicated that all *McPPAD* genes, as well as that encoding the peritrophin McPM1, were expressed in the midgut of all stages of feeding and molting larvae. Additionally, *McPPAD1*, *2*, and *3* genes as well as *McPM1* (Chapter 4) were expressed in foregut, hindgut, Malpighian tubules, fat body, tracheae and integument in addition to the midgut, albeit at slightly different levels. More cycles (40 vs 28) were required to obtain a visible product from *McPPAD3* transcripts, indicating that this gene was expressed at a lower level (Figure 6.3).

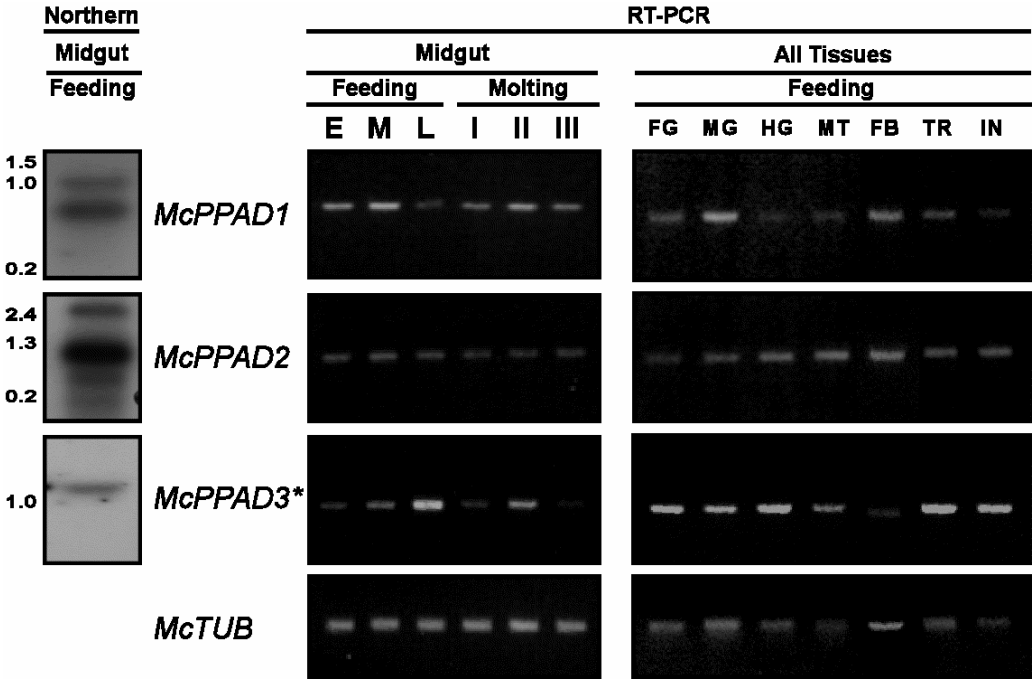


Figure 6.3 Expression of genes encoding *Mamestra configurata* proteins with peritrophin-A domains (McPPADs). Total RNA from the midgut (MG) of 4th instar feeding larvae was used in the northern blot analysis. The sizes of the molecular weight markers (kb) are shown in the left hand margin of the northern blot panel. Total RNA from the midgut of early (E), middle (M) and late (L) stage feeding 4th instar larvae and Phases I, II and III molting larvae (from 4th to 5th instar) was used for RT-PCR analysis with 28 or 40(*) cycles. Gene expression patterns were examined in foregut (FG), midgut (MG), hindgut (HG), Malpighian tubules (MT), fat body (FB), tracheae (TR) and integument (IN). Amplification of the *M. configurata* tubulin (*McTUB*) gene was used as a control in the RT-PCR to ensure equivalent RNA template in all samples.

6.3.1.4 Localization

The predicted molecular weight of McPPAD1 was 23 kDa; however, rMcPPAD1 expressed in *E. coli* BL21 (DE3) migrated as a ~32 kDa protein on SDS-PAGE gels (Figure 6.4A). A peptide derived antiserum was used to localize McPPAD2 because expression of this protein was unsuccessful in all *E. coli* strains used. rMcPPAD3 expressed in *E. coli* Rosetta 2 (DE3) migrated at ~37 kDa in SDS-PAGE gels which was larger than its predicted 28 kDa size (Figure 6.4A).The anti-McPPAD1 antiserum reacted only with rMcPPAD1, while the anti-McPPAD3 antiserum reacted only with rMcPPAD3 (Figure 6.4B). The anti-McPPAD2 did not react with either rMcPPAD1 or rMcPPAD3, but did recognize the peptides used for the immunization in an ELISA assay (data not shown).

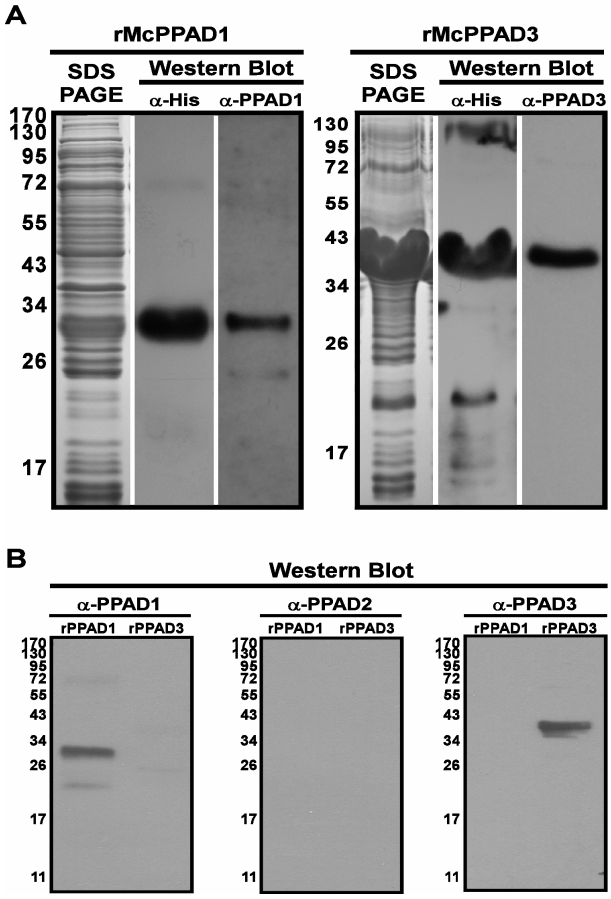


Figure 6.4 Western blot analysis of recombinant *Mamestra configurata* peritrophin A domain protein 1 (rMcPPAD1) and peritrophin A domain protein 3 (rMcPPAD3) with α-PPAD1-3 antisera. Blots were probed with either antihexahistidine (α-His) antibody (A) or α-PPAD1, α-PPAD2 and α-PPAD3 antisera (A and B).

The anti-rMcPPAD1 antiserum reacted with a protein of >170 kDa, and also with smaller proteins of approximately 95 and 62 kDa in the PM (Figure 6.5A, B) confirming the liquid chromatography tandem-mass spectrometry (LC-MS/MS) analysis that indicated association of McPPAD1 protein with PM (Chapter 4, Table 4.2). The same antiserum reacted also with a protein of 39 kDa in the foregut and hindgut, and with a 113 kDa protein in Malpighian tubules. In midgut samples, the antiserum reacted with a major protein of 113 kDa and three minor proteins of 95, 72 and 62 kDa. When proteins were extracted from midgut epithelium after removal of the basal lamina with dispase, the antiserum reacted with proteins of 113, 83 and 32 kDa in the epithelial fraction alone, whereas several weak bands were observed with the basal laminal fraction (Figure 6.5B). A BLAST search of the 14,000 midgut ESTs with McPPAD1 did not detect any other proteins that had regions of similarity extensive enough to suggest that antiserum would cross-react with another protein; therefore, the higher molecular bands, at least in the midgut samples, are likely a complex of proteins that include McPPAD1.

The peptide derived McPPAD2 antiserum detected a single protein of approximately 17 kDa only in Malpighian tubules (Figure 6.5C). This was larger than the 12 kDa predicted for this protein, although *O*-linked glycosylation (two sites predicted) would increase its molecular weight. The two peptides used to generate this antiserum were used to search the midgut EST database using BLAST. One of the peptides was unique, while the second showed 70% identity to a protein represented by a single partial EST from the midgut cDNA library; however, BLAST analysis did not reveal any orthologs in other species. There may be potential for antiserum cross-reactivity given that the 43 kDa protein/region in the midgut sample reacts very weakly with the McPPAD2 antiserum; however, the protein in Malpighian tubules reacts very strongly.

The anti-rMcPPAD3 antiserum reacted with a protein of approximately 37 kDa in the foregut, midgut and integument (Figure 6.5D). In addition, a 20 kDa protein was detected in the midgut. Longer exposures also revealed a band approximately 37 kDa in tracheae and the hindgut (data not shown). The peptide derived antiserum detected the same proteins as the antiserum derived from the recombinant protein, verifying the western blot results (data not shown).

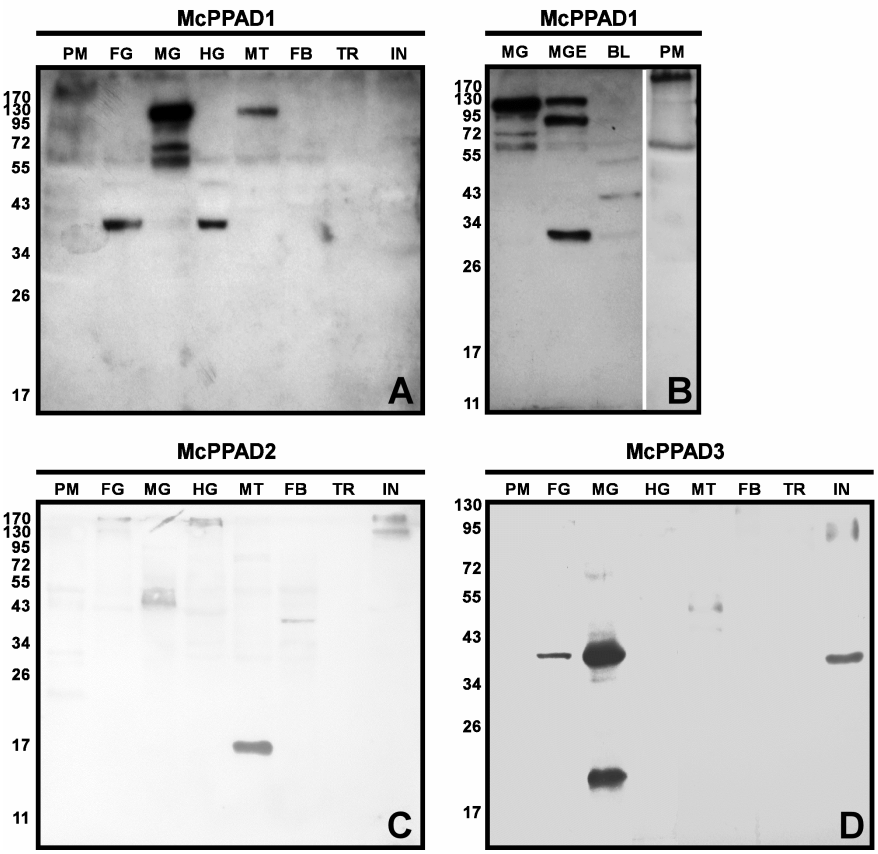


Figure 6.5 Western blot analysis showing the location of *Mamestra configurata* proteins with peritrophin A domains (McPPADs). A, B: McPPAD1, C: McPPAD2, D: McPPAD3. Basal lamina (BL), fat body (FB), foregut (FG), hindgut (HG), integument (IN), Malpighian tubules (MT), midgut (MG), midgut epithelium (MGE), peritrophic matrix (PM) and tracheae (TR) were used. The sizes of the molecular weight markers (kDa) are shown in the left hand margin of each panel.

6.3.2 Identification and Classification of Lepidopteran Midgut Proteins with Peritrophin Domains

Search of midgut EST databases for proteins containing PADs, PBDs and PCDs identified nine *M. configurata* (including McPPAD1-4) and five *B. mori* PPADs; however, no proteins were found that had PBDs or PCDs. The *B. mori* PPADs were named according to the nomenclature used for the probable *M. configurata* orthologs. The BLAST search of the NCBI database using *M. configurata* PPADs including peritrophins identified 32 PPADs (Appendix F) from 5 lepidopteran families: Bombycidae (*B. mori*), Noctuidae (*H. armigera*, *M. configurata*, *S. exigua*, *S. frugiperda*, *T. ni*), Plutellidae (*P. xylostella*), Tortricidae (*E. postvittana*), and Pyralidae (*L. sticticalis*).

Four types of PPADs were identified based on the organization of the PADs and MDs (Figure 6.6). The “simple PPADs” contain one to three PADs and may contain small MDs. The “binary PPADs” have two PADs separated by a MD. The ‘complex PPADs’ contain multiple PADs interspersed with MDs with a carboxy terminal PAD containing eight cysteines. The “repetitive PPADs” are composed of multiple (6-19) tandem PADs. Among these protein groups, binary, complex, and repetitive PPADs include only the peritrophins and their obvious orthologs; however, simple PPADs may contain proteins associated with other tissues in addition to PM.

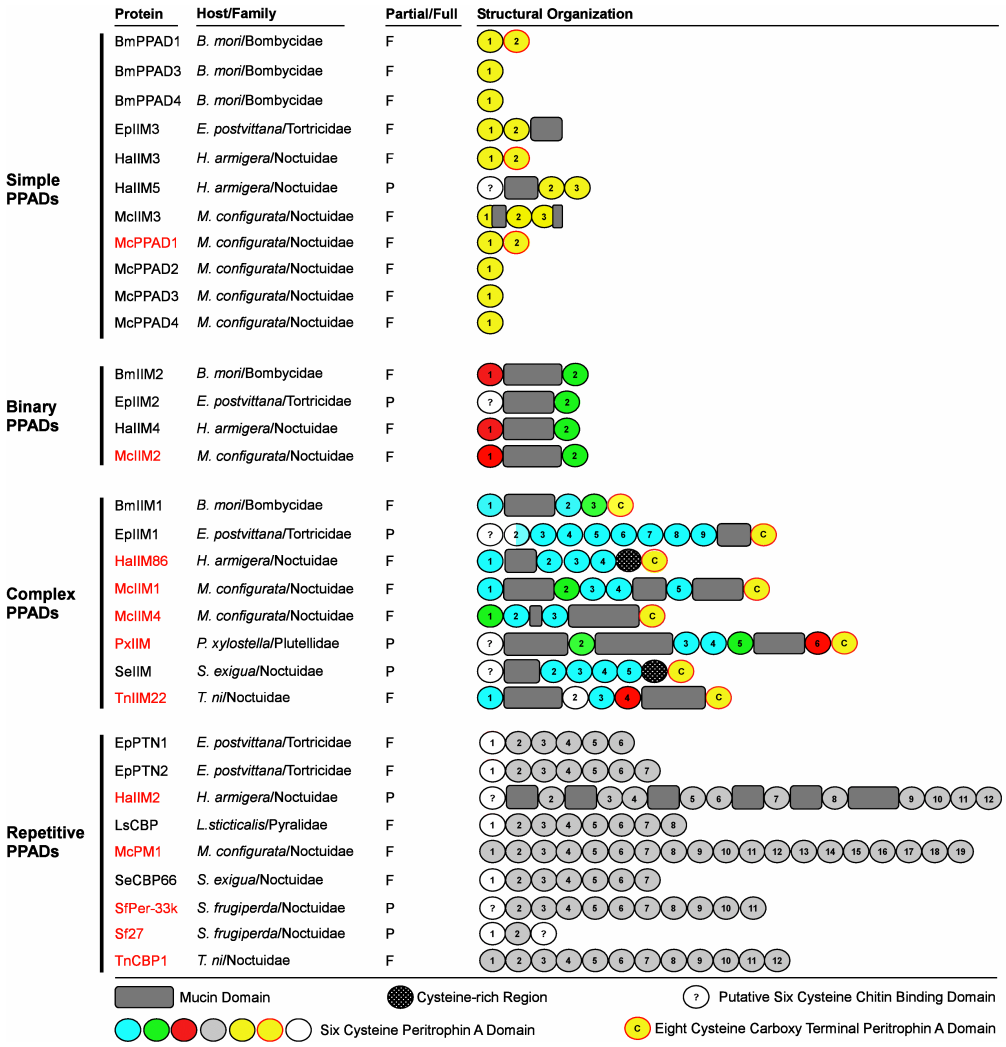


Figure 6.6 Structural organization of lepidopteran proteins with peritrophin-A domains (PPADs). The colour scheme relates to the PAD clades in Figure 6.7 and to the amino acid sequences in Appendix F. When full length PPADs were not available, the putative amino terminal PAD is denoted with a “?” and subsequent PADs numbered sequentially beginning with 2. The names of the PPADs known to be associated with the peritrophic matrix are in red lettering.

6.3.3 Phylogenetic Analysis of PADs and Development of an Evolutionary Model

Phylogenetic analysis using 150 PADs (Appendix G) from the 32 PPADs (Appendix F) revealed 2 main groups (Figure 6.7). One group contained PADs found in the simple, binary and complex PPADs, whereas the second group contained the majority of PADs present in the repetitive PPADs. The phylogenetic relationship between the PADs (Figure 6.7) in combination with the organization of elements within the PPADs (Figure 6.6) was used to develop an evolutionary model (Figure 6.8). In the model, an ancestral gene encoding a single PAD duplicates (Event I), with each gene giving rise to a separate lineage. In Lineage 1, a gene duplication (Event II) occurs with one gene giving rise to the Class I simple PPADs represented by BmPPAD3-4, EpIIM3, McIIM3, McPPAD2-4 and HaIIM5. These simple PPADs have so far not been demonstrated to be associated with the PM. This group also contains McIIM3 which has unique hybrid PAD-MDs. The PAD in the second gene is duplicated and then the gene itself is duplicated (Event III). One gene gives rise to the Class II simple peritrophins (BmPPAD1, HaIIM3 and McPPAD1) while the carboxy terminal domain in the other acquires two additional cysteines. Subsequently, the amino terminal PAD is duplicated to form an intermediate that is the progenitor of the binary and complex PPADs. The binary PPADs (BmIIM2, EpIIM2, HaIIM4 and McIIM2) arise after the eight cysteine terminal PAD is lost and a MD is acquired. The complex PPADs evolve from sequential duplication of the amino and internal PADs while the eight cysteine carboxy terminal PAD remains fixed. Lineage 2 leads to the repetitive PPADs (EpPTN1, EpPTN2, HaIIM2, LsCBP, McPM1, SeCBP66, SfPer-33K and TnCBP1). Here, the PAD undergoes an initial duplication with the carboxy terminal PAD subsequently undergoing multiple, rapid duplications, as indicated by their high degree of similarity to one another (Figure 6.7). The most repetitive representative of this type, McPM1, has 19 PADs.

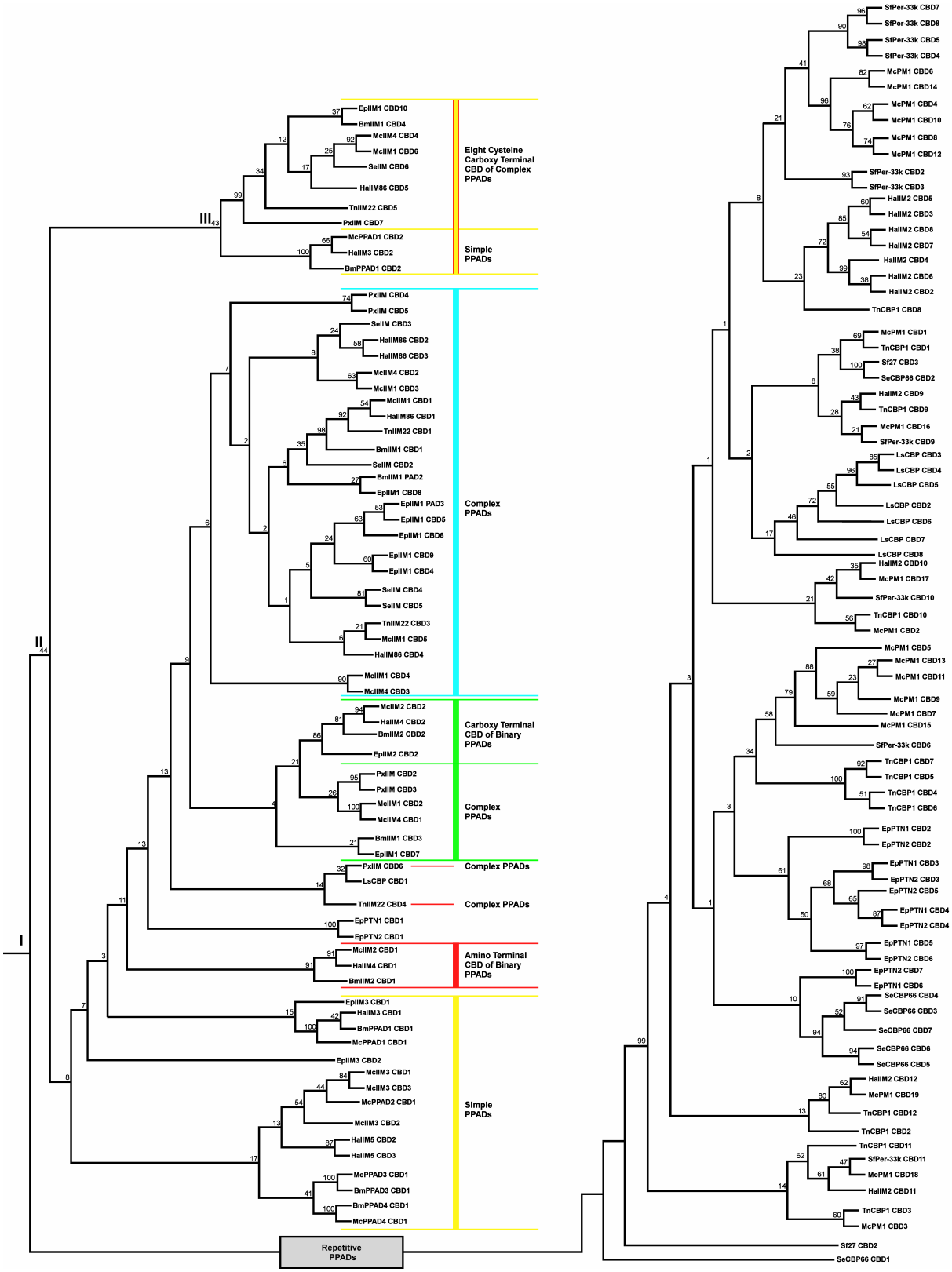


Figure 6.7 (legend is on the next page)

Figure 6.7 Phylogenetic analysis of chitin binding domains from lepidopteran proteins with peritrophin-A domains (PPADs). The following proteins were used in the analysis: *Mamestra configurata* McIIM1 (GenBank accession number, AY057052), McIIM2 (FJ670567), McIIM3 (FJ670568), McIIM4 (FJ670569), McPM1 (AY277403), McPPAD1 (GU596430), McPPAD2 (GU596431), McPPAD3 (GU596432), McPAD4 (GU596433); *Plutella xylostella* PxIIM (AF545582); *Trichoplusia ni* TnIIM22 (AF000606), TnCBP1 (AAR06265); *Helicoverpa armigera* HaIIM86 (ABW98670), HaIIM2 (EU325543), HaIIM3 (EU325564), HaIIM4 (EU325565), HaIIM5 (EU449965); *Epiphyas postvittana* EpIIM1 (EV809288), EpIIM2 (EV809261), EpIIM3 (EV809248), EpPTN1 (EV809224), EpPTN2 (EV806871); *Spodoptera frugiperda* SfPer-33k (contig assembled from AAS89976 + contig 32 provided by Dr. W.R. Terra + Sf2M02530-5-1 from SpodoBase), Sf27 (provided by Dr. W.R. Terra); *Bombyx mori* BmIIM1 (BGIBMGA000185-PA), BmIIM2 (BGIBMGA001480-PA), BmPAD1 (BGIBMGA009641-PA), BmPPAD3 (BGIBMGA003270-PA), BmPPAD4 (BGIBMGA011851-PA); *S. exigua* SeCBP66 (ABW98673), SeIIM (EU047712); *Loxostege sticticalis* LsCBP (ACQ65651). The colour scheme denoting the distinct groups of PADs corresponds to the structural organization (Figure 6.6) and the evolutionary model (Figure 6.8). The symbols I, II and III correspond to major evolutionary events denoted in Figure 6.8.

6.3.4 Acquisition of Mucin Domains in IIMs

The model predicts four independent routes for IIM evolution. Three of these lead to the simple, binary and complex PPADs within Lineage 1. IIMs among the simple PPADs evolve by acquiring small, distinct MDs (McIIM3, EpIIM3 and HaIIM5). McIIM3, which has two hybrid PAD-MDs, appears unique as an ortholog was not found in any of the other lepidopteran PPADs examined. All the binary PPADs are IIMs, and their MDs were acquired soon after PAD duplication as all the full length members contain similar amino acid repeats (e.g., PTT; Appendix I). With few exceptions, the MDs in the complex peritrophins are distinct and have different amino acid repeat elements (Appendix I) indicating that they were acquired independently. An IIM, HaIIM2, was also found among the repetitive PPADs and therefore represents a fourth independent route for mucin evolution, in this case within Lineage 2. HaIIM2 contains multiple MDs interspersed among the PADs (Figure 6.8). The first five MDs in this peritrophin have TAAP amino acid repeats indicating they have a common origin. As the PADs in this protein are closely related, a module containing a MD and one or two PADs duplicated repeatedly. The last MD has a different amino acid repeat, suggesting it was acquired independently.

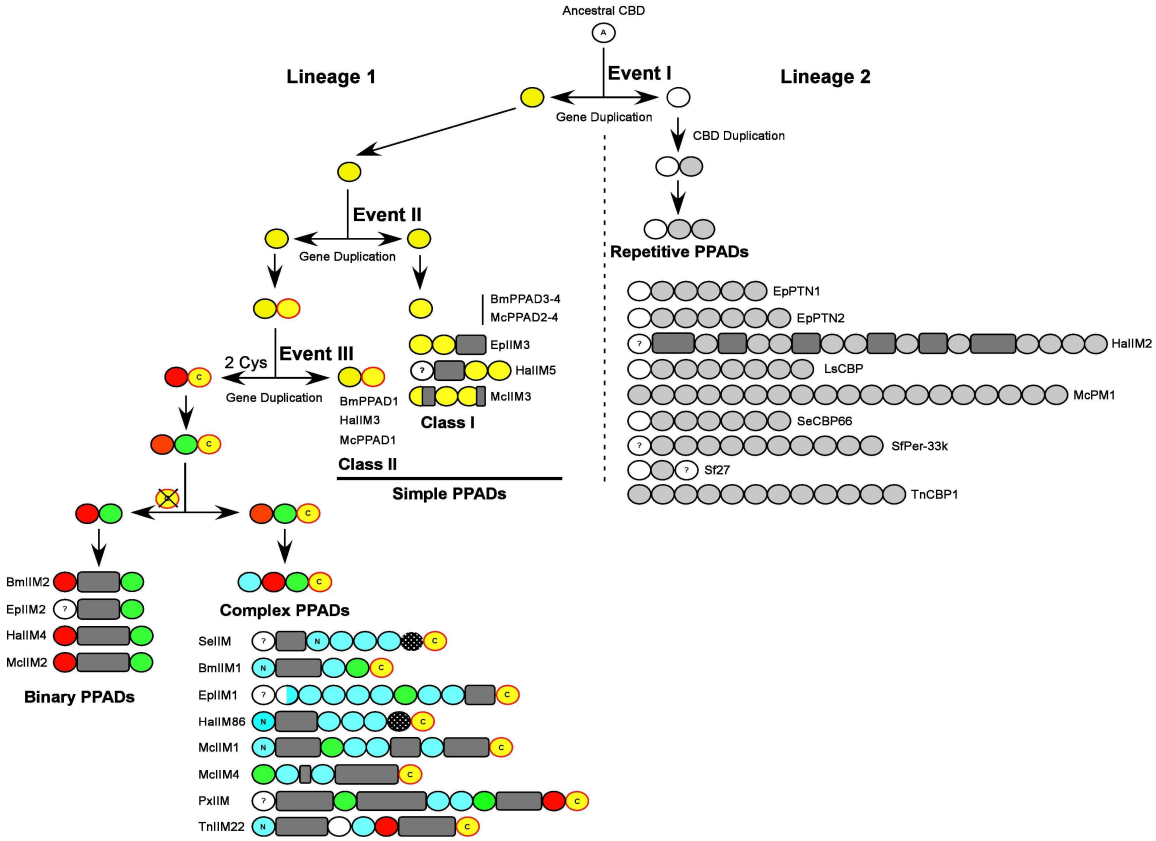


Figure 6.8 Model showing the proposed evolution of lepidopteran proteins with peritrophin-A domains (PPADs) based on phylogenetic analysis of PADs in combination with the structural organization of the domains. Major gene duplication events are indicated with Roman numerals and these correspond to branch points in Figure 6.7. The colour scheme and nomenclature relate to the PAD groups established in Figure 6.6.

6.3.5 PAD Modeling in support of the Phylogenetic Analysis

Protein fold recognition analysis using tachycitin as a guide structure (Figure 6.9A) was able to model the two PADs of McPPAD1 (Figure 6.9C). The structures were similar to tachycitin which contains three antiparallel β-sheets at the amino terminus and a hevein-like fold with two antiparallel β-sheets followed by an α-helix at the carboxy terminus. C2 and C6 were predicted to be sufficiently proximate to form a disulphide bridge as were C4 and C5. The program was unable to model the amino terminal region containing C1. Modeling of the PADs of McPPAD1 revealed a genuine chitin binding pocket (chitin binding pocket 1) into which project the R-groups of the first aromatic amino acid between C2 and C3, as well as those of the aromatic amino acid between C4 and C5. The R-group of the second aromatic amino acid between C2 and C3 projects into a second region which may also have a second putative chitin

binding pocket (chitin binding pocket 2) (Figure 6.9C). McPPAD2 lacked both aromatic amino acids between C2 and C3 and the aromatic amino acid between C4 and C5 was one amino acid further downstream than the conserved position. McPPAD3 had only a single aromatic amino acid between C2 and C3 in addition to that between C4 and C5. The software was unable to model the PADs from McPPAD2 and McPPAD3 since the identity values were less than 20% compared to the tachycitin CBD (Figure 6.9B).

The phylogenetic analysis indicated that the carboxy terminal CBD2 from simple peritrophins (typically with a six cysteine PAD) and the eight cysteine carboxy terminal PAD from the complex peritrophins were related (Figure 6.7). The typical six cysteine PAD in simple peritrophins had the consensus register $CX_{15}CX_5CX_9CX_{12}CX_7C$ (Table 6.4), while the eight cysteine PAD has two additional cysteines located four amino acids downstream of C1 (denoted C1' and four amino acids upstream of C6 (denoted C6'), in addition to the regular six cysteine register. The consensus motif for the latter PAD was $CX_3CX_{9\text{-}10}CX_5CX_9CX_{12}CX_3CX_3C$. To confirm that the eight cysteine PAD was a genuine PAD and could have shared a common ancestor with the six cysteine PAD, a structural model of the eight cysteine carboxy terminal PAD in McIIM1 was also developed based on the tertiary structure of tachycitin. The amino terminal PAD (CBD1) of McIIM1, which has six cysteines was also analyzed. This also revealed that both six and eight cysteine PADs of McIIM1 have three antiparallel β-sheets at the amino terminus and a hevein-like fold with two antiparallel β-sheets followed by an α-helical region at the carboxy terminus as in the case of CBD2 of McPPAD1 (Figure 6.9C). In the modeling of six or eight cysteine PADs, disulphide bonds were predicted between C2-C6 and C4-C5 as in McPPAD1; however, C1 and C3 were not sufficiently proximate to form a bond corresponding to that in tachycitin (Figure 6.9C). Therefore, the structure of this region in the PAD must differ slightly from that of tachycitin to allow the critical C1-C3 bond to form. Interestingly, the model of the eight cysteine PAD predicted that C1' could form a bond with C3 and therefore the structure of this region may be more similar to that of tachycitin. Modeling of the McIIM1 PADs also revealed that a putative chitin binding pocket 1 contains two aromatic amino acids (first aromatic amino acid between C2 and C3 and the aromatic amino acid between C4 and C5) and a putative chitin binding pocket 2 that contains a single aromatic amino acid (the second aromatic amino acid between C2 and C3) (Figure 6.9C).

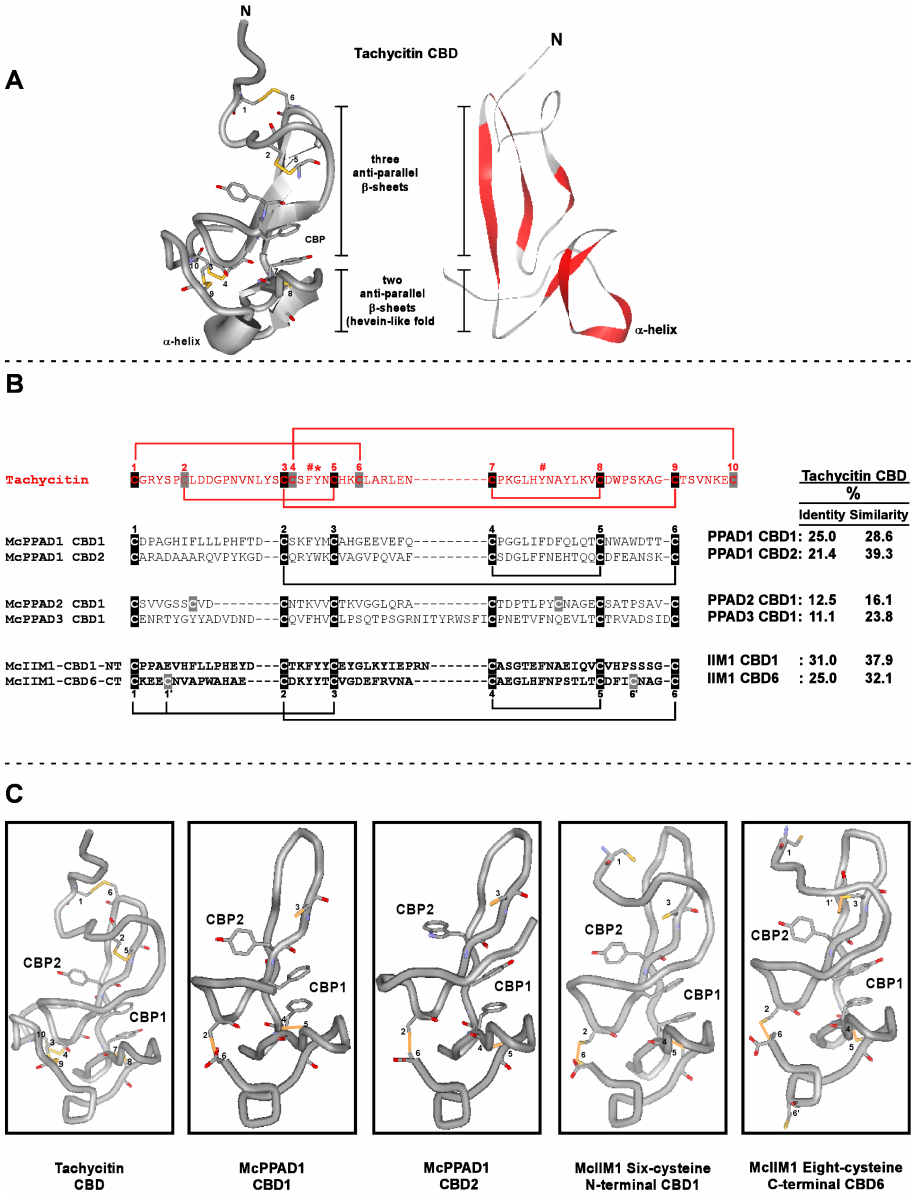


Figure 6.9 Modeling of *Mamestra configurata* peritrophin A domains according to tachycitin chitin binding domain. A: Cartoon showing the three antiparallel β-sheets in the amino terminus and the hevein-like chitin binding domain (CBD) fold at the carboxy terminus of tertiary model of tachycitin. B: Alignment of the tachycitin CBD (GenBank accession number, D85756) with *M. configurata* McPPAD1, McPPAD2, McPPAD3, and McIIM1 amino (NT) and carboxy terminal (CT) CBDs. Conserved (white on black background) and unique (white on grey background) cysteines are numbered with putative disulphide bridges indicated by lines. Aromatic amino acids projecting into the putative chitin binding pockets shown as (#) or (*). C: Tertiary structure models of tachycitin (Protein Data Bank code 1DQC) showing secondary structure features, and *M. configurata* McPPAD1 CBD1 and CBD2, and McIIM1 amino terminal CBD1 and carboxy terminal CBD6. Numbers indicate cysteines forming disulphide bonds (orange lines). Aromatic amino acids projecting into the putative chitin binding pocket (CBP) are shown.

6.4 Discussion

6.4.1 Characterization of McPPADs

6.4.1.1 Structure and composition

PADs are the most common type of cysteine-containing CBDs in lepidopterans and coleopterans, while PBD and PCD have been reported only from dipterans (Tellam et al., 1999). The previously established consensus for PADs was $C^1X_{13-20}C^2X_{5-6}C^3X_{9-19}C^4X_{10-14}C^5XC_{4-14}C^6$ (Tellam et al., 1999). This can now be expanded to $C^1X_{9-20}C^2X_{5-6}C^3X_{9-19}C^4X_{10-17}C^5X_{4-14}C^6$ when all the PADs from lepidopteran PPADs are included (Table 6.4). The PADs from McPPAD1 and McPPAD3 are within these limits.

The regions flanking the McPPAD1 and McPPAD3 PADs are hydrophilic. Regions separating the McPM1 and McIIM PADs, referred to as spacers, also contain a high proportion of charged amino acids interspersed with prolines (Shi et al., 2004). This arrangement was proposed to generate tightly coiled structures that would bring PADs into close proximity to one another along the chitin bundle (Shi et al., 2004). The regions separating the PADs in CfGasp, an ortholog of Obstructor-C (Obst-C) which has three PADs and is synthesized by cuticular tissues in *D. melanogaster*, from the lepidopteran *Choristoneura fumiferana*, are also hydrophilic (Nisole et al., 2010). The spacer regions may not have an analogous function in proteins containing a single or a few PADs; however, charged amino acid clusters may allow for interaction with other proteins or molecules. Furthermore, McPPAD1 was predicted to have an *N*-glycosylation site. The di-*N*-acetylglucosamine in the core glycan could mimic chitin and therefore be recognized by a CBD of another peritrophin, which would serve to promote the formation of protein complexes.

There are only a few reports of simple PPADs and this is the first report from Lepidoptera of PPADs with a single PAD, other than chitinases and CDAs, though such proteins have been described in dipterans (Eisemann et al., 2001), coleopterans (Venancio et al., 2009) and siphonapterans (Gaines et al., 2003). Obst proteins in *D. melanogaster* and their orthologs in *D. pseudo-obscura, An. gambiae and Apis mellifera* contain three PADs (Behr and Hoch, 2005). Recently, an ortholog of *Obst-C* was identified in the lepidopterans *C. fumiferana* and *B. mori* (Nisole et al., 2010) and in the coleopteran *T. castaneum* (Jasrapuria et al., 2010), suggesting that this type of PAD protein may be quite common.

6.4.1.2 Expression

Genes encoding PPADs, McPM1 and McPPAD1-3, are expressed in a range of tissues, specifically the foregut, midgut, hindgut, Malpighian tubules, fat body, tracheae and integument. The expression of genes encoding PPADs, including peritrophins, was once thought to be restricted to tissues synthesizing PM, namely the midgut and cardia. Indeed, *McIIM4*, which encodes an IIM containing both PAD and MDs, is expressed only in the midgut (Chapter 4). However, *PPAD* genes are now known to be expressed in other tissues including *L. cuprina* hindgut (Tellam et al., 2003), *Mayetiola destructor* salivary glands and fat body (Mittapalli et al., 2007), *An. gambiae* salivary glands (Dimopoulos et al., 1998), *C. felis* hindgut and Malpighian tubules (Gaines et al., 2003), and *D. melanogaster* ovaries (Barry et al., 1999). In *D. melanogaster*, the *Obst* family of genes each encode a protein with three PADs (*Obst-A* to *Obst-J*) and members are expressed in tissues associated with cuticle including integument, tracheae, foregut and hindgut, as well as in midgut (Behr and Hoch, 2005). *Obst-C*, previously referred to as *Gasp*, is also expressed in embryonic tracheae (Barry et al., 1999). Expression of *CfGasp* from *C. fumiferana* is low in larval midgut, tracheae and ovaries but high in the epidermis and head where cuticle is a major component (Nisole et al., 2010). Recently, the entire genomic complement of *T. castaneum PPAD* genes, referred to as *Type 2 CBD* or *ChtCBD2* genes, was catalogued and divided into one of two groups based on their expression patterns. Genes expressed exclusively in the midgut of feeding larvae were classified as “Peritrophic Matrix Protein” genes, while those expressed in nonmidgut tissues throughout larval development were referred to as “Cuticular Proteins Analogous to Peritrophins” genes (Jasrapuria et al., 2010), though localization studies of the proteins were not conducted in either case. The wide distribution in the expression of *McPPAD* genes in *M. configurata* indicates that such a clear distinction cannot be made for this species and possibly other lepidopterans. It should be noted that although *McPM1* and *McPPAD1-3* transcripts were found at similar levels in a wide range of tissues, the corresponding protein was not necessarily detected in all these tissues indicating that a mechanism exists to control the translation of their mRNAs.

Table 6.4 Arrangement of cysteines within peritrophin-A domains (PADs) from lepidopteran proteins with PADs.

Species	Protein	Cysteine Consensus [1]					
B. mori	BmIIM1	C^1X_{13-15}	C^2X_5	C^3X_{9-11}	C^4X_{12-13}	C^5X_7	C^6
	BmIIM2	C^1X_{15-18}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6
	BmPPAD1	C^1X_{15}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6
	BmPPAD3	C^1X_{14}	C^2X_5	C^3X_{18}	C^4X_{12}	C^5X_8	C^6
	BmPPAD4	C^1X_{14}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_8	C^6
C. fumiferana	CfGASP	C^1X_{12-20}	C^2X_5	C^3X_9	C^4X_{12-17}	C^5X_{7-8}	C^6
E. postvittana	EpIIM1	C^1X_{14-15}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6
	EpIIM2	C^1X_{15}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6
	EpIIM3	C^1X_{14-15}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6
	EpPTN1	C^1X_{14}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6
	EpPTN2	C^1X_{14}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6
H. armigera	HaIIM86	C^1X_{13-15}	C^2X_5	C^3X_{9-11}	C^4X_{12}	C^5X_7	C^6
	HaIIM2	C^1X_{14}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6
	HaIIM3	C^1X_{15}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6
	HaIIM4	C^1X_{15-18}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6
	HaIIM5	C^1X_{13}	C^2X_5	C^3X_{9-10}	C^4X_{12-13}	C^5X_7	C^6
L. sticticalis	LsCBP	C^1X_{14}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6
M. configurata	McIIM1	C^1X_{13-15}	C^2X_5	C^3X_{9-11}	C^4X_{12}	C^5X_{7-8}	C^6
	McIIM2	C^1X_{15-18}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6
	McIIM3	C^1X_{13-15}	C^2X_5	C^3X_{13-15}	C^4X_{12}	C^5X_5	C^6
	McIIM4	C^1X_{13-15}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_{7-8}	C^6
	McPM1	C^1X_{14-15}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6
	McPPAD1	C^1X_{15}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6
	McPPAD2	C^1X_9	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6
	McPPAD3	C^1X_{14}	C^2X_5	C^3X_{18}	C^4X_{12}	C^5X_8	C^6
	McPPAD4	C^1X_{14}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_8	C^6
P. xylostella	PxIIM	C^1X_{14-15}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6
S. exigua	SeCBP66	C^1X_{13-14}	C^2X_5	C^3X_9	C^4X_{10-12}	C^5X_7	C^6
	SeIIM	C^1X_{15}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6

Table 6.4 (continued)

Species	Protein	Cysteine Consensus [1]					
S. frugiperda	Per-33k	C^1X_{14}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6
	Sf27	C^1X_{13-14}	C^2X_5	C^3X_9	C^4X_{10-12}	C^5X_7	C^6
T. ni	TnIIM22	C^1X_{13-15}	C^2X_5	C^3X_{9-11}	C^4X_{12}	C^5X_7	C^6
	CBP1	C^1X_{14-15}	C^2X_5	C^3X_9	C^4X_{12}	C^5X_7	C^6

[1] The general arrangement of cysteines in a PAD for lepidopteran PPADs is $\mathbf{C^1}X_{9-20}\mathbf{C^2}X_5\mathbf{C^3}X_{9-18}\mathbf{C^4}X_{10-17}\mathbf{C^5}X_{5-8}\mathbf{C^6}$.

McPM1 and *McPPAD1*, which encode peritrophins, are expressed in the midgut of all stages of feeding and molting larvae, including premolt larvae in which feeding has ceased and the old PM has begun to disintegrate. This observation indicates that peritrophin gene expression is not dependent on PM synthesis, which is in accordance with reports for *IIM* genes. Like most lepidopterans, *M. configurata* larvae have a Type I PM and its production is almost continuous. Continuous production of peritrophins may be necessary as they are proposed to bind to pre-existing immature PM during the process of PM maturation and to stabilize the PM network by interlocking the chitin fibrils (Tellam et al., 2000). Such proteins might also accumulate in secretory vesicles prior to delivery into the lumen when the new PM begins to form as in mosquitoes (Devenport et al., 2004; 2005).

Genes encoding cuticle associated PAD proteins, McPPAD3 and CfGasp (Nisole et al., 2010) are expressed continually throughout larval development, though *CfGasp* expression is low during the intermolt and peaks during the molt (Nisole et al., 2010).

6.4.1.3 Localization

McPPAD proteins are located in tissues other than PM. Although the PAD was first identified in peritrophins (Tellam et al., 1999), it is now very clear that it is not exclusive to them. PADs have also been found in other proteins such as a *S. frugiperda* midgut chitinase (Bolognesi et al., 2005), proteins associated with *C. felis* Malpighian tubules, hindgut, rectum and tracheae (Gaines et al., 2003) and *An. gambiae* epidermis (He et al., 2007), and in baculovirus proteins, for example, Autrographa californica multiple nucleopolyhedrovirus open reading frame 145 (ORF145) and ORF150 (Lapointe et al., 2004). Therefore, the definition of

peritrophins as integral PM proteins which have at least one CBD and can be extracted from the PM by strong denaturants (Tellam et al., 1999) should remain.

McPPAD1 is found in the digestive tract and Malpighian tubules. In the PM, McPPAD1 was detected in a >170 kDa band which was much larger than its predicted molecular weight of 23 kDa. Peritrophins may migrate aberrantly on SDS-PAGE gels due to their glycosylation and elevated proline and cysteine contents (Shen and Jacobs-Lorena, 1998; Shen et al., 1999). The difference between the predicted and the observed migration behaviour for McPPAD1 was too great to have been caused by such modifications, as it has only a single predicted *N*-linked glycosylation site and a proline content of 5.0%. Rather, it may be that McPPAD1 is a member of a protein complex. McPPAD1 may be covalently linked to other proteins in this complex since the SDS and NaCl in the extraction buffer should have disrupted ionic interactions, while the β-mercaptoethanol would disrupt disulphide bonds in the CBDs and prevent their binding to the core of *N*-linked glycans as noted above. This complex could help to anchor the PM to the midgut epithelium as peritrophin CBDs might recognize *N*-linked microvillar glycoproteins present in the glycocalyx as proposed by Ferreira et al. (2007).

In midgut tissue, McPPAD1 is also associated with a higher molecular protein band. Treatment with dispase released two lower molecular weight proteins that were also recognized by the McPPAD1 antiserum providing further evidence that it may be part of a protein complex. Furthermore, one of the proteins was 32 kDa, which was similar in molecular weight to rMcPPAD1. Peritrophin-33k, a protein associated with the PM of *S. frugiperda*, was also detected in midgut microvilli (Ferreira et al., 2007), though it is not an ortholog of McPPAD1.

Although they contain a PAD, neither McPPAD2 nor McPPAD3 localize to the PM. The presence of McPPAD2 in Malpighian tubules is unexpected since these organs are composed of a single layer of epithelial cells situated on the inner side of a basal lamina and do not contain chitin (Gillott, 2005). In *C. felis*, the PPAD, PL1, is mainly in Malpighian tubules. A chitin binding assay showed that PL1 does not bind to chitin *in vitro*, suggesting that the PAD might bind other types of sugar moieties (Gaines et al., 2003), which may also be the case for McPPAD2. McPPAD3 is found mostly in foregut, midgut and integument and to a lesser extent in tracheae and hindgut. A common feature of the lining of the foregut, hindgut, tracheae as well as cuticular exoskeleton, is the presence of chitin. The prevalent chitin binding motif in cuticular proteins is the “R&R consensus sequence” first identified by Rebers and Riddiford (1988);

however, this is not present in McPPAD3 suggesting that it is a nontypical cuticular protein. Similarly, *Gasp/Obst-C* from *D. melanogaster* as well as *CPAP1* and *CPAP3* from *T. castaneum* are expressed primarily in cuticular tissues, but the corresponding proteins do not have an R&R consensus sequence (Barry et al., 1999; Jasrapuria et al., 2010; Nisole et al., 2010).

6.4.2 Development of an Evolutionary Model for Lepidopteran PPADs

It is worth noting that the peritrophins with PADs could be considered a subclass of PPADs, and all the lepidopteran peritrophins, except *M. configurata* CBD3P identified in Chapter 4, have a PAD. Therefore, the four classes identified in the current study represent also the peritrophin classes. Indeed, the binary, complex and repetitive PPADs include only peritrophins and their orthologs. However, simple PPADs (except McPPAD1 and its ortholog BmPPAD1) are associated with other tissues as demonstrated for McPPAD2 and McPPAD3 in the current study.

Previous analysis of CBDs from invertebrate chitinases and peritrophins indicated that the two groups evolved from a common ancestor (Shen and Jacobs-Lorena, 1999). The current study extended this analysis and focused on PADs from all known lepidopteran PPADs. The evolutionary model developed in the current study proposes that the PADs in the repetititive PPADs and the Class I simple PPADs constitute the most ancient PAD lineages, while the eight cysteine carboxy terminal PADs were derived from the PADs that gave rise to Class II simple PPADs. Further, it is proposed that the loss of the eight cysteine carboxy terminal PADs preceded the evolution of binary PPADs while the complex PPADs evolved with the retention of the eight cysteine carboxy terminal PAD.

The four major classes of PPADs differentiate regardless of whether the majority rule or extended majority rules were used in the phylogenetic analysis. In other words, the clades contain CBDs from the same structural class of proteins. Although the bootstrap values for many branches of the tree generated using the extended majority rule were lower than 50%, this approach was necessary to tease out details within each class and most PPADs conform to this model. EpPTN1 CBD1, EpPTN2 CBD1, LsCBP CBD1, PxIIM CBD6 and TnIIM22 CBD4 cluster between the PADs from the complex PPADs and the amino terminal CBD of the binary PPADs. Since these PADs are found among peritrophins from the repetitive and complex peritrophins, any similarity may simply be fortuitous.

The model reveals four independent routes for IIM evolution; thus, each group contains a single type of mucin. It was proposed that IIMs evolved from gastrointestinal mucins by acquiring PADs (Terra, 2001); however, the organization within the larger set of IIMs used in the current study does not appear to support this. Rather, peritrophins containing only PADs likely acquired varied MDs on several different occasions. Regardless of the primary sequence, all MDs show high potential for *O*-linked glycosylation, a feature proposed to be fundamental to their role in protecting the PM from mechanical, chemical and proteolytic damage (Wang and Granados, 1997a).

6.4.3 PAD Modeling in support of the Phylogenetic Analysis

The PADs of McPPAD1 as well as the amino and carboxy terminal PADs of McIIM1 were successfully modeled using tachycitin CBD as a guide, as were the PADs from *D. melanogaster* Obstructor (Obst) proteins, a group of secreted chitin binding proteins associated with cuticle forming tissues (Behr and Hoch, 2005). Tachycitin was reported to have three antiparallel β-sheets at the amino terminus as well as a hevein-like fold consisting of two antiparallel β-sheets and followed by an α-helical region at the carboxy terminus (Suetake et al., 2000). This general structural organization was also predicted for each of the modeled PADs including the eight cysteine carboxy terminal PAD in McIIM1. Thus, both phylogenetic and protein recognition analyses suggest that six and eight cysteine PADs are closely related.The PADs from McPPAD2 or McPPAD3 could not be modeled using this approach as they have low (<20%) sequence identity with the tachycitin CBD. Furthermore, McPPAD2 has additional cysteines between C1 and C2 and C5 and C6 within the PAD cysteine register. Although the PAD of McPPAD3 conforms to the consensus, it contains nine more amino acids between C3 and C4 than the analagous region in tachycitin and this may have contributed to the inability of the software to model the PAD.

The disulphide bridges formed between the cysteines are important for CBD stability. The tachycitin CBD has 10 cysteines ($C^1X_5C^2X_{11}C^3C^4X_4C^5X_2C^6X_6C^7X_{12}C^8X_7C^9X_6C^{10}$) that form 5 disulphide bridges between C1-C6, C2-C5, C3-C9, C4-C10 and C7-C8 (Kawabata et al., 1996; Suetake et al., 2000) (Figure 6.9). Six of these cysteines (C1, C3, C5, C7, C8 and C9) correspond to the six cysteine register of PAD and were proposed to form three disulphide bridges between C1-C3, C2-C6 and C4-C5 (Shen and Jacobs-Lorena, 1999). The models showed

that bonds could form between C2 and C6 and C4 and C5 in both six and eight cysteine PADs which agrees with this proposal; however, C1 and C3 were not sufficiently close to form a bond in McIIM1 six and eight cysteine PADs (Figure 6.9). In contrast, C1' was predicted to form a bond with C3 in the eight cysteine carboxy terminal PAD, suggesting the structure of this region may be different between the two PAD isoforms or between both PADs and tachycitin. Interestingly, replacement of C1 and C3 with alanine was reported to eliminate chitin binding activity of an *Ae. aegypti* PAD, indicating that involvement of these cysteines in disulphide bridge formation is essential for chitin binding activity (Shao et al., 2005). Thus, this finding also supports the hypothesis that the corresponding amino terminal regions in six and eight cysteine PAD isoforms or tachycitin may be different.

Binding of chitin by PADs is mediated by hydrophobic interaction between aromatic amino acids and β-(1, 4)-*N*-acetyl-D-glucosamine (GlcNAc) (Shao et al., 2005). The arrangement of aromatic amino acids within carbohydrate binding domains is related to their function (Boraston et al., 2004). Thus, carbohydrate binding domains with aromatic amino acids lying planar along the surface recognize crystalline polysaccharides, while aromatic amino acids projecting into an extended groove are associated with enzymatic functions. In a third type, aromatic amino acids reside in a chitin binding pocket and form a "sandwich" platform that allows interaction with the α and β faces of the GlcNAc pyranose ring (Boraston et al., 2004). Suetake et al. (2000) examined the structural similarities between the hevein-like region of tachycitin and hevein-32. For hevein-32, oligosaccharide interaction was shown to be mediated by specific amino acids including W^{21} and W^{23} and aided by S^{19}. Superimposition of the tachycitin structure showed that these corresponded to Y^{49}, V^{52} and N^{47}, respectively; however, the conserved aromatic amino acid in the PADs corresponds to Y^{46}. Furthermore, the amino terminal region consisting of three antiparallel β-sheets contributes a second aromatic R-group from F^{27} to the same chitin binding pocket (CBP1). Another conserved aromatic amino acid corresponding to Y^{28} of tachycitin projects into a second pocket; however, there is as yet no direct evidence that this amino acid is involved in chitin binding.

6.4.4 Potential Role of PPADs

Although all the lepidopteran peritrophins (except McCBP3) are PPADs, all the PPADs are not peritrophins; thus, some of them may have functions other than contributing to PM

structure and integrity. A specific role of certain PPADs may be difficult to predict due to their wide tissue distribution as shown for McPPAD1-3. Interestingly, McPPAD1 is associated with the tissues in the digestive tract and Malpighian tubules. Peritrophin-15 from *Chrysomya bezziana* and *L. cuprina* contains a single CBD and binds to chitin at a low stoichiometry, leading to the hypothesis that such proteins with 1-2 CBDs cap the end of individual chitin chains and protect the fibrils from exochitinases (Wijffels et al., 2001).

McPPAD2 is an integral component of Malpighian tubules, which lack chitin. However, it contains an atypical PAD that lacks certain conserved aromatic amino acids required for chitin binding suggesting that it may have evolved an alternate function.

McPPAD3 is found in chitinous tissues such as integument, foregut, hindgut and tracheae. *Gasp* genes are expressed primarily in the epidermis and upregulated during the molt in *C. fumiferana* suggesting they may be involved in cuticle formation (Nisole et al., 2010). Similarly, Obst proteins were proposed to be required for chitin packaging during cuticle differentiation (Behr and Hoch, 2005). Tracheal proteins with PADs were proposed to be involved in gas exchange and protection from entry of viruses, or parasites (Barry et al., 1999). By analogy, McPPAD3 may also serve in one of these capacities.

In conclusion, the current chapter showed that genes encoding proteins with one or two PADs are expressed in multiple tissues without necessarily being translated into proteins. Such a wide distribution of gene expression has not been reported previously for the genes encoding PPADs including peritrophins. Additionally, PADs are not unique to peritrophins or midgut associated proteins in lepidopterans. Structural organization together with phylogenetic analysis of PADs allowed development of an evolutionary model and a classification of PPADs including peritrophins (simple, binary, complex and repetitive). Peritrophins containing MDs were found among each of the classes and evolved independently on several occasions. The typical PAD with a six cysteine register was also closely related to a second PAD form, containing two additional cysteines, which is found exclusively at the carboxy terminus of the complex peritrophins. Modeling of both PAD forms based on tachycitin tertiary structure revealed that they also have similar hevein-type folds with three antiparallel β-sheets at the amino terminus and two β-sheets followed by an α-helical region at the carboxy terminus. The organization of the disulphide bonds formed between C2-C6 and C4-C5 were as predicted; however, the C1 and C3 were not predicted to form a bond in the typical six cysteine PAD whereas C1' was predicted

to form a bond with C3 in the eight cysteine PAD. Finally, both PAD forms were predicted to form one or two separate chitin binding pockets.

The previous chapters examined various aspects of PM biology, including morphology, chitin synthesis and protein composition. The following chapters will focus on determining the role specific proteins play in PM function and whether this information can be used to develop novel pest control strategies. Specifically, Chapter 7 will examine the interaction of IIMs with the baculoviral metalloprotease, enhancin, and Chapter 8 the use of RNA interference to disrupt gene expression.

7. ROLE OF INSECT INTESTINAL MUCINS IN BACULOVIRUS INFECTIONS

7.1 Introduction

Insect intestinal mucins (IIMs) are common components of the insect peritrophic matrices (PMs) and contain one or more *O*-glycosylated regions that are rich in threonines, serines and prolines (Wang and Granados, 1997a). The IIMs are proposed to protect the PM from physical or chemical damage and protease degradation, and the midgut epithelial cells from pathogen invasion, while lubricating the passage of food materials (Terra, 2001).

The family *Baculoviridae* includes arthropod specific viruses with large, circular, covalently closed, double stranded DNA genomes between 80 and 180 kbp. The family is comprised of four genera, *Alphabaculovirus* [formerly known as *Nucleopolyhedrovirus* (NPV) infecting lepidopteran hosts], *Betabaculovirus* [formerly known as *Granulovirus* (GV)], *Gammabaculovirus* (hymenopteran specific NPV) and *Deltabaculovirus* (dipteran specific NPV) (Jehle et al., 2006). Baculoviruses can have two virion phenotypes, budded virus (BV) and occlusion derived virion (ODV). BVs are formed early in the infection cycle when nucleocapsids bud through the host cell membrane and acquire an envelope. BVs are responsible for cell-to-cell transmission of the virus within the host insect. Late in the infection cycle, nucleocapsids are enveloped by membranes formed within the infected host nucleus and these virions are subsequently occluded within proteinaceous crystals or occlusion bodies (OBs). These ODVs are infectious upon ingestion by a susceptible host, but ODV must first cross the peritrophic matrix (PM) to infect midgut epithelial cells.

Early studies showed that cofeeding of GV OBs with NPV OBs significantly increased the infectivity and virulence of the NPV for its natural host even in the absence of any actual GV infection (Tanada, 1959; Yamamoto and Tanada 1978; Zhu et al., 1989). In the case of *Trichoplusia ni granulovirus* (TnGV), this enhancement of NPV infectivity was linked to degradation of the host insect PM and increased PM permeability by a metalloprotease, enhancin, that was associated with the OBs (Lepore et al. 1996; Peng et al. 1999). Subsequently, TnGV enhancin was shown to specifically degrade *Trichoplusia ni* insect intestinal mucin (TnIIM) (Wang and Granados 1997a; b). TnGV *enhancin* was the first baculovirus *enhancin* gene to be sequenced (Hashimoto et al., 1991). In addition, *enhancin* genes were found in the genomes of numerous betabaculoviruses (reviewed in Galloway et al., 2005), as well as in a few Group II alphabaculoviruses (reviewed in Galloway et al., 2005; Thiem, 2009), including

Mamestra configurata nucleopolyhedrovirus (MacoNPV-A and MacoNPV-B) (Li et al. 2002a; b; 2003). The enhancins so far investigated are produced during the late phase of infection (Bischoff and Slavicek, 1997) and are packaged within the OBs in the case of betabaculoviruses or appear to be ODV proteins, in the case of alphabaculoviruses. Alphabaculovirus enhancins have a transmembrane domain and Lymantria dispar multiple nucleopolyhedrovirus (LdMNPV) enhancin was localized to the ODV envelope (Slavicek and Popham, 2005). In contrast, enhancins from betabaculoviruses lack a transmembrane domain and are incorporated into OBs (Gallo et al., 1991; Roelvink et al., 1995). These differences suggest that the impact of enhancin activity on the virus infection process may be somewhat different in the two baculovirus genera.

As indicated above, the viral enhancing activity of purified enhancin has been well documented in betabaculoviruses (Gallo et al., 1991; Wang et al., 1994; Wang and Granados, 1998). Further evidence for this was provided by experiments that used a recombinant Autographa californica multiple nucleopolyhedrovirus (AcMNPV) expressing TnGV enhancin (Lepore et al., 1996). Similar approaches have been used for alphabaculoviruses; for example, the LD_{50} value of a recombinant AcMNPV expressing MacoNPV enhancin was 4.4 times lower than that of wild type AcMNPV in second instar *T. ni* larvae (Li et al., 2003). As well, deletion of *enhancin* genes in LdMNPV decreased viral potency confirming that it enhances activity of this virus (Bischoff and Slavicek, 1997; Popham et al., 2001). However, only a few studies have attempted to address the mode of action of enhancins. Early studies pointed to an interaction between betabaculovirus enhancins and the PM (Derksen and Granados, 1988) and enhancin purified from TnGV OBs degraded TnIIM *in vivo* and *in vitro* (Wang and Granados 1997a; b). As a consequence of this interaction, the permeability of *Pseudaletia unipuncta* and *T. ni* PMs to AcMNPV virions and blue dextran was increased (Peng et al., 1999). Only indirect evidence is available for the mode of action of alphabaculovirus enhancin. Experiments with MacoNPV-A showed that unprocessed or partially processed forms of McIIM1 were degraded after *per os* challenge of *Mamestra configurata* larvae (Shi et al., 2004).

To date, no detailed studies of the activity of alphabaculovirus enhancins have been undertaken in terms of their interaction with host insect PM. In the current study, the interaction between MacoNPV and the McIIMs that were characterized in Chapter 5 are examined. Specifically, the impact of MacoNPV enhancin, as a representative of alphabaculovirus enhancins, on different IIM types is investigated.

7.2 Material and Methods

7.2.1 Virus Isolates

MacoNPV-A strain 90/2, was originally isolated from a field collected, single larval cadaver (Li et al., 2002a). MacoNPV-A OBs were purified from infected 4th instar *M. configurata* larvae as previously described (Erlandson, 1990). Similarly, OBs of AcMNPV strain E-2 and an AcMNPV recombinant, AcMNPV-enMP2, expressing MacoNPV-A *enhancin* under the control of its native promoter (Li et al., 2003) were purified from infected 4th instar *T. ni* larvae.

7.2.2 Oral Exposure Assays

One hundred 4th instar larvae were fed 5.0x10^5 MacoNPV-A OBs/larva applied to 4 mm diameter *B. napus* leaf disks as previously described (Erlandson et al., 2007). In parallel, 120 4th instar larvae fed *B. napus* leaf disks without virus were used as uninfected controls. PMs from 20 infected and 20 uninfected larvae were dissected 2, 4, 6, 12, and 24 hours post inoculation (hpi). Loosely and strongly associated PM proteins were extracted from the samples as described in Chapter 2.

7.2.3 In vitro *Exposure Assays*

Two sets of *in vitro* experiments were performed. In the first set, western blot analysis was used to determine the integrity of McIIM4 in PMs incubated in the presence or absence of alkali solubilized MacoNPV-A OBs in order to confirm the *in vivo* data (degradation of McIIM4 upon *per os* inoculation). Twelve PMs from middle stage 4th instar feeding larvae were dissected and rinsed in ddH_2O to remove the food bolus. The PMs were transferred to a 1.5 ml microcentrifuge tube with 7.5x10^6 MacoNPV-A OBs and Ringer's solution (153 mM NaCl, 2.68 mM KCl, 1.36 mM $CaCl_2 \cdot 2H_2O$) adjusted to a final volume of 100 µl. The suspension was mixed thoroughly using a vortex, and Na_2CO_3 and $ZnCl_2$ were added to a final concentration of 20 mM and 10 mM, respectively, and the mixture was incubated at 28ºC in a shaker (225 rpm) for 1 h. An additional aliquot of 7.5x10^6 MacoNPV-A OBs in 20 mM Na_2CO_3 and 10 mM $ZnCl_2$ was added, mixed with a vortex and incubated overnight at 28ºC in a shaker. The following day, the mixture was centrifuged at 14,000 g and the PM pellet washed three times in 200 µl of Ringer's

solution. In parallel, PMs without virus were treated and incubated in the presence $ZnCl_2$ and Na_2CO_3 using the same conditions as with the virus treatments.

The second set of experiments was designed to test whether the degradation of McIIM4 was due specifically to MacoNPV enhancin. To do this, $1.5x10^7$ OBs of AcMNPV-wt, AcMNPV-enMP2 or MacoNPV-A were tested using the protocol described above. In addition, the two enhancin coding viruses (MacoNPV-A and AcMNPV-enMP2) were tested in the presence of a metalloprotease inhibitor, EDTA, at a final concentration of 10 mM. McIIM2 integrity was also examined in these experiments with the same conditions as those used for McIIM4.

7.2.4 Protein Separation and Western Blot Analysis

Protein concentrations were determined using the Bradford assay and for *in vitro* experiments equal amounts of proteins were loaded onto one dimensional SDS-PAGE gels, separated by electrophoresis and electroblotted onto nitrocellulose membranes as described in Chapter 2. For *in vivo* experiments, protein from an equal number of PMs was loaded. The strongly associated PM protein fractions were separated by SDS-PAGE, and McIIM2 and McIIM4 content was examined by western blot analysis as described in Chapter 2.

7.3 Results

7.3.1 Oral Exposure Assays

SDS-PAGE profiles of strongly associated PM protein fractions from control and MacoNPV-A challenged larvae were similar at 2, 6, 12 and 24 hpi. Conversely, the protein profile at 4 hpi from the PMs of MacoNPV-A challenged larvae (Figure 7.1B) was different from that of uninfected control larvae (Figure 7.1A) as additional bands between 34 and 95 kDa were present. A band of 32 kDa, demonstrated earlier to be predominantly soybean lectin, GmLe1, from the diet (Chapter 4), gradually increased with time in PM protein samples from both uninfected and MacoNPV-A challenged larvae (Figure 7.1A, B).

The anti-McIIM2 antiserum reacted strongly with a protein of approximately 130 kDa and weakly with two higher molecular weight proteins (≥170 kDa) in the PMs of both uninfected (Figure 7.1C) and MacoNPV-A challenged (Figure 7.1D) larvae at all time points. The anti-McIIM4 antiserum reacted strongly with a protein of approximately 170 kDa, which is likely the

fully processed form of McIIM4, in PM protein samples from uninfected (Figure 7.1E) and MacoNPV-A challenged (Figure 7.1F) larvae at all time points. The amount of this protein increased concomitantly with the amount of time the larvae were feeding. The major difference was that the anti-McIIM4 antiserum reacted with multiple proteins between 25 and 86 kDa in the PM sample from MacoNPV-A challenged larvae at 4 hpi (Figure 7.1F). In the control samples, the antiserum also reacted with proteins of 70 and 88 kDa that may represent the unglycosylated and partially glycosylated forms of McIIM4. These proteins were not observed in the samples from challenged larvae. A 54 kDa protein, which could be a product of differential splicing of the McIIM4 mRNA (Chapter 5), was present in both control and MacoNPV-A challenged larvae at 12 and 24 hpi. Finally, a protein of approximately 46 kDa was present in the samples from MacoNPV-A challenged larvae at 12 and 24 hpi. Several other proteins that were common to both MacoNPV-A challenged and uninfected larvae cross-reacted with the antiserum. These included the 32 kDa protein containing the soybean lectin, which was found to gradually increase with time.

7.3.2 In vitro *Exposure Assays*

SDS-PAGE protein profiles of PMs treated with alkali solubilized MacoNPV-A OBs also exhibited several bands that were not present in the profile from untreated PMs (Figure 7.2A). Eleven proteins were common to both PMs, while six additional proteins appeared in the MacoNPV-A treated samples (Figure 7.2A). Western blot analysis with the anti-McIIM4 antiserum revealed numerous proteins in the samples from treated PMs ranging from 24 to 52 kDa (Figure 7.2B) of which five were of the same size (25, 34, 39, 42 and 52 kDa) as those detected in the the PM of *per os* challenged larvae (Figure 7.1F).

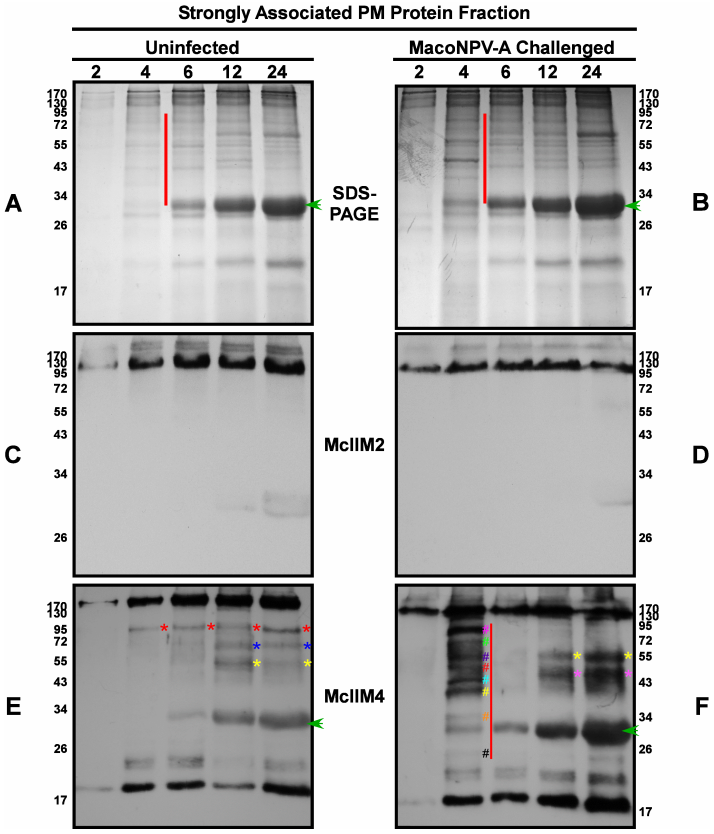


Figure 7.1 Analysis of strongly associated proteins from the peritrophic matrix (PM) of *Mamestra configurata* larvae *per os* challenged with MacoNPV-A. SDS-PAGE analyses of strongly associated proteins from the peritrophic matrix (PM) of uninfected control (A) or MacoNPV-A challenged (B) 4th instar *M. configurata* larvae. The region showing putative degradation products of PM proteins from MacoNPV-A challenged larvae (red line in B) and corresponding region in uninfected control larvae (red line in A) are indicated. Western blot analyses of McIIM2 and McIIM4 from the PM of uninfected (C, E) and MacoNPV-A challenged (D, F) 4th instar larvae at 2, 4, 6, 12, and 24 hpi are shown. Putative unglycosylated (blue or yellow *) and partially glycosylated (red *) forms of McIIM4 are indicated in E and F. The putative degradation products (red line in F) which comprised eight bands with approximate molecular weights of 25 (black # symbol), 34 (orange), 39 (yellow), 42 (light blue), 48 (red), 52 (purple), 70 (green), 88 (pink) kDa of McIIM4 from the PMs of MacoNPV-A challenged larvae are indicated. A protein of 46 kDa (purple *) found only in samples from MacoNPV-A challenged larvae is also indicated in F. The location of soybean lectin (green arrow) and the sizes of molecular weight markers (kDa) are shown in the left or right hand margins.

To examine whether the degradation of McIIM4 was due to the activity of enhancin, a second set of *in vitro* experiments was conducted using two additional viruses, AcMNPV-wt, which naturally lacks an *enhancin* gene, and the recombinant virus, AcMNPV-enMP2, which expresses the MacoNPV-A *enhancin.* Incubation of the PM with MacoNPV-A released five

proteins (25, 34, 39, 42 and 52 kDa) (Figure 7.3B) that corresponded to those observed in the previous *in vitro* (Figure 7.2B) and *in vivo* (Figure 7.1F) experiments. Three of these proteins (25, 42 and 52 kDa) were also released from PMs treated with AcMNPV-enMP2. In the presence of the metalloprotease inhibitor, EDTA, these proteins were not released after MacoNPV-A and AcMNPV-enMP2 treatment, indicating that McIIM4 degradation was due to MacoNPV-A enhancin. Both untreated control PMs and those treated with AcMNPV-wt had two of these proteins (42 and 52 kDa). Interestingly, the 54 kDa protein was present in the control, AcMNPV, MacoNPV-A and AcMNPV-enMP2 treatments, but absent in the presence EDTA, indicating that it may derive from processing of McIIM4 by endogenous metalloproteases rather than from a differentially spliced mRNA. No degradation products derived from McIIM2 were detected in any of the virus treatments, in accordance with the result of the *per os* feeding experiment (Figure 7.3A).

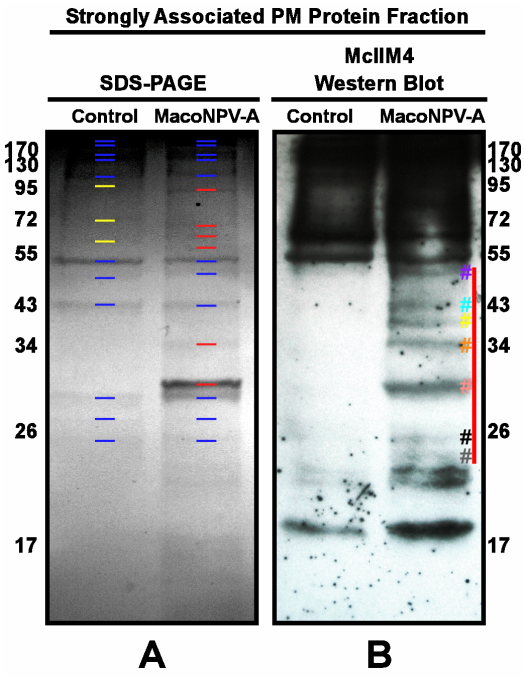


Figure 7.2 Analysis of *Mamestra configurata* McIIM4 from the peritrophic matrix (PM) incubated with alkali solubilized MacoNPV-A OBs. SDS-PAGE analysis of strongly associated protein fractions from PMs incubated without (control) or with alkali solubilized MacoNPV-A OBs (A) and corresponding western blot analysis of McIIM4 (B) are shown. Horizontal lines in A indicate the common (blue) or unique bands (red or yellow). The vertical red line in B indicates the putative degradation products of McIIM4, which comprised seven proteins of approximately 24 (gray # symbol), 25 (black), 30 (pink), 34 (orange), 39 (yellow), 42 (light blue) and 52 kDa (purple). Molecular weight markers (kDa) are shown in the left or right hand margin.

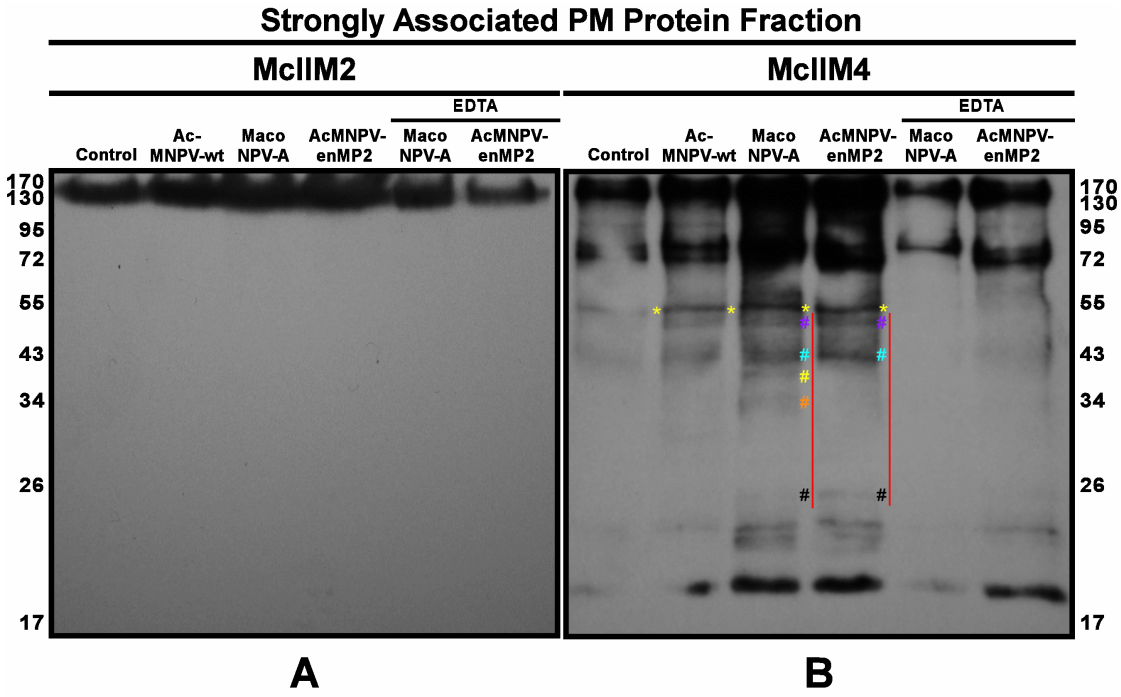


Figure 7.3 Analysis of *Mamestra configurata* McIIM2 and McIIM4 from the peritrophic matrix (PM) incubated with alkali solubilized AcMNPV-wt, MacoNPV-A or AcMNPV-enMP2 OBs. Western blot analysis of McIIM2 (A) and McIIM4 (B) from strongly associated protein fractions of peritrophic matrices (PMs) incubated without any virus (control) or with alkali solubilized virus OBs is shown. Viruses encoding enhancin (MacoNPV-A and AcMNPV-enMP2) were also tested in the presence of the metalloprotease inhibitor, EDTA. The vertical red line in B indicates the putative degradation products of McIIM4, which comprised five bands of approximately 25 (black # symbol), 34 (orange), 39 (yellow), 42 (light blue) and 52 kDa (purple). The 54 kDa protein (yellow *) is also indicated. Molecular weight markers (kDa) are shown in the left or right hand margin.

7.4 Discussion

Western blot analysis using anti-McIIM4 antiserum revealed multiple bands, representing the degradation products of McIIM4, between 25 and 86 kDa in the PM protein samples from *M. configurata* larvae at 4 hpi. Shi et al. (2004) also reported that McIIM1 was degraded 4 h after MacoNPV-A challenge, suggesting that both McIIM1 and McIIM4 are degraded soon after inoculation and that this may allow the virions to pass through the PM to the midgut epithelium. Interestingly, the fully glycosylated form represented by a band of approximately 170 kDa, was present at all time points post-inoculation suggesting a major factor affecting McIIM degradation is glycosylation. Shi et al. (2004) also reported the degradation of unprocessed or partially processed forms of McIIM1 after MacoNPV-A oral inoculation, while fully glycosylated forms remained intact. In contrast, McIIM2 was not affected after *per os* MacoNPV-A challenge. As

discussed in Chapter 5, IIM glycosylation refers to *O*-glycosylation occurring at serines or threonines in the MDs. MDs comprise 32.7, 49.4, and 42.4% of the total number of amino acids in McIIM1, McIIM2, and McIIM4, respectively. In addition, the percentage of glycosylated serines or threonines within the MDs is 23.7-40, 70.1 and 23.5-39.7% for McIIM1, McIIM2 and McIIM4, respectively. Therefore, the largest cumulative MD contribution occurs in McIIM2 and the degree of *O*-glycosylation is the greatest for the McIIM2 MD. These factors may contribute to the resistance of McIIM2 to degradation by MacoNPV-A enhancin. Indeed, high *O*-glycosylation potential is a common feature of all binary IIMs that were described as proteins with a large central mucin domain flanked by CBDs (Chapter 6). Presumably, glycosylation physically protects the protein from proteolytic enzymes, such as enhancin (Wang and Granados, 1997a). The smaller size of MDs relative to the size of the IIM they are found in, and lower degrees of glycosylation, may be the reason for the degradation of McIIM4 or McIIM1, both of which are complex IIMs as reported in Chapter 6. Interestingly, TnIIM, which was also degraded after ingestion of baculoviral enhancin or treatment of the PM (Wang and Granados, 1997a; b), is also a complex IIM, suggesting that IIMs of this class are sensitive to proteolysis.

Mucin synthesis and glycosylation are dynamic processes (Karlsson et al., 2000) and different McIIM4 forms are present in the PM, including the fully glycosylated form (≥170 kDa) and smaller forms that may represent partially glycosylated (≥88 kDa) or unprocessed forms (54 and 70 kDa) (Chapter 5). The 70 and 88 kDa proteins were absent in MacoNPV-A challenged larvae at 6-24 hpi suggesting that some virus remains and that these highly susceptible forms continue to be degraded well after the initial challenge. Shi et al. (2004) reported that all forms of McIIM1 were recovered in the PM by 24 hpi. Competition between IIM synthesis and degradation by the baculovirus may quickly change the protein patterns. This is likely dependent on the dose of OBs as was demonstrated for TnIIM degradation by baculoviral enhancin (Wang and Granados, 1997a). The current study and that of Shi et al. (2004) used relatively high virus dosages, $2.0x10^5$ and $5.0x10^5$ OBs/larva, respectively, indicating these two IIMs may have different structural features affecting their integrity and persistency against MacoNPV-A challenge.

The *in vitro* experiments were designed and carried out to maximize the number of MacoNPV ODVs interacting directly with the PM and to test whether MacoNPV-A enhancin was responsible for McIIM4 degradation. In these experiments, multiple bands between 24-52

kDa representing the degradation products of McIIM4 were detected in samples extracted from PMs incubated with alkali solubilized MacoNPV-A OBs. Among these, proteins of 25, 34, 39, 42, and 52 kDa were of the same size as those found in the PMs of MacoNPV-A *per os* challenged larvae, indicating that there are consistent McIIM4 degradation products upon exposure to MacoNPV-A ODVs both *in vivo* and *in vitro*. Treatment of PM with the AcMNPV-enMP2 resulted in a similar pattern of McIIM4 degradation as that observed for MacoNPV-A; however, 34 and 39 kDa bands were not detected. Although the same dose was used for both viruses, the amount of recombinant enhancin in AcMNPV-enMP2 ODV may be lower than that in the native MacoNPV-A ODV, resulting in a lower degree of degradation. Furthermore, no degradation was detected when either enhancin-encoding virus was incubated with PMs in the presence of EDTA, suggesting that the degradation of McIIM4 is due to the activity of MacoNPV-A enhancin. This finding is the first demonstration of IIM degradation due to an alphabaculovirus enhancin; however, this has been reported for betabaculovirus enhancins. For example, Wang and Granados (1997a) reported *in vitro* degradation of recombinant TnIIM by TnGV enhancin and the degradation pattern was similar to the McIIM4 degradation pattern obtained *in vitro* in the current study. As well, EDTA was shown to inhibit degradation of TnIIM, confirming that enhancin is a metalloprotease (Lepore et al., 1996). Similarly, feeding of *T. ni* neonates with purified TnGV enhancin caused TnIIM to be degraded in the midgut (Wang and Granados, 1997a). A *Bacillus thuringiensis* enhancin-like enzyme was also shown to degrade the IIM of *Helicoverpa armigera* and severely disrupt the PM (Fang et al., 2009).

The *in vitro* experiments also confirmed *in vivo* findings that McIIM2 is not targeted, providing the first report of an IIM that is not degraded by baculoviral enhancins. The high glycosylation potential suggests that this class of IIMs may be specifically involved in protecting the PM from proteolytic enzymes. This aspect should be examined further in an *in vitro* system using recombinant IIMs, proteases and *O*-glycosidases.

Interestingly, *in vitro* treatment of PM with AcMNPV-wt also produced a weaker pattern of McIIM4 degradation with two proteins (42 and 52 kDa) in common with the MacoNPV-A treatment; however, most of the other proteins (25, 34, and 39 kDa) seen upon MacoNPV-A challenge *in vitro* (Figure 7.3B) or *in vivo* were missing. It is noteworthy that serine proteases were shown to be associated with baculovirus OBs (Maskel and DiCapua, 1988) and this may explain the degradation of McIIM4 observed with AcMNPV-wt ODV *in vitro*. It may also be

possible that some degradation occurred during the dissection as evidenced by the appearance of weak McIIM4 degradation products in the untreated samples. The PM is likely to lose its integrity to a certain degree and may be less resistant to the proteolytic factors in the bolus, which is a protease rich environment. Since these experiments aimed to elucidate the role of enhancin in McIIM4 degradation, the PM was not treated with a protease inhibitor cocktail during the dissection.

Although alphabaculovirus and betabaculovirus enhancins target PM proteins, they have some structural differences that may affect their modes of action. For example, TnGV enhancin lacks a transmembrane domain, is mostly associated with the OBs and comprises approximately 5% of the total OB protein (Roelvink et al., 1995). Therefore, when the GV OBs dissolve in the midgut, enhancin release may have a dramatic effect on IIMs. In contrast, transmembrane domains are present at the carboxy terminus of all alphabaculovirus enhancins, including MacoNPV-A enhancin, and they localize to the ODV envelope (Slavicek and Popham, 2005). Therefore, the impact of alphabaculovirus enhancin on IIMs will be only at the sites where the ODVs interact with the PM, suggesting that the degradation of the target IIMs may be more subtle in contrast to the prominent degradation caused by betabaculovirus enhancins.

In conclusion, the current study revealed that McIIM4 is degraded by MacoNPV-A enhancin, while McIIM2 is not. Understanding the roles of IIMs and the features that contribute to these may aid the development of new insect control strategies. Baculovirus enhancins are promising in this regard because these proteins can degrade the PM and promote infection by baculoviruses. In the following chapter, another promising approach for insect control and a valuable tool for understanding the roles of proteins, namely, gene silencing by RNA interference, is examined *in vitro* and *in vivo* for genes encoding specific PM components.

8. TARGETING GENES INVOLVED IN PERITROPHIC MATRIX FORMATION BY RNA INTERFERENCE

8.1 Introduction

One drawback to peritrophic matrix (PM) research is the lack of tools to investigate the precise role of PM components, such as peritrophins or chitin synthase and chitinolytic enzymes, in PM function or maintenance of integrity. Although many studies have been undertaken to identify these components, most could only infer function based on sequence features, gene expression patterns or protein localization. PM functional genomics studies would be greatly enhanced by an efficient and specific method to disrupt gene expression. Traditional approaches, including genetic screens and knockout technology, have been important in this regard; however, their use is often limited to model organisms or require a long time to properly identify and characterize the mutants. A new approach for specific gene disruption was developed based on the Nobel Prize winning discovery by Fire et al. (1998). They found that when specific double strand RNA (dsRNA) was injected into the nematode, *Caenorhabditis elegans*, the corresponding mRNA disappeared from both the somatic cells of the organism and those of its F1 progeny (Fire et al., 1998). This specific gene silencing was referred to as "RNA interference" (RNAi). RNAi has quickly become a widely used tool to knock down and analyze the function of genes in many organisms, including mammals, plants, fungi and nematodes. Furthermore, the dsRNA may be delivered by injection, feeding or soaking the organism in dsRNA solution (Fire et al., 1998; Tomoyasu et al., 2008). The RNAi pathway is triggered by introduction of dsRNA into a cell, upon which the RNase III enzyme, dicer, cleaves the dsRNA into fragments of 21-23 nucleotides each called small-interfering RNAs (siRNAs). These siRNAs remain as dsRNA duplexes with very short 3' overhangs (Elbashir et al., 2001) and direct the RNA induced silencing complex (RISC) to degrade target mRNAs complementary to the siRNAs (Martinez et al., 2002).

The first *in vivo* demonstration of RNAi in insects was from the dipteran, *Drosophila melanogaster*, after injection of dsRNA into embryos (Kennerdell and Carthew, 1998). This was followed by *in vitro* cell culture demonstrations in *D. melanogaster* S2 cells (Hammond et al., 2000) and many other studies that used soaking of embryos (Eaton et al., 2002) or injection of adults (Kato et al., 2006). RNAi has also been successful in species from other orders, including coleopterans and hymenopterans, by injection of dsRNA into embryos (Brown et al., 1999; Beye

et al., 2002), larvae (Tomoyasu and Denell, 2004), pupae (Bucher et al., 2002) or adults (Amdam et al., 2003) or by feeding larvae with *in vitro* amplified or plant expressed dsRNA (Baum et al., 2007). There are several recent successful demonstrations of RNAi in lepidopterans in which the dsRNA is delivered by feeding larvae with *in vitro* amplified (Turner et al., 2006) or plant expressed dsRNA (Mao et al., 2007) or by injection of dsRNA into larvae (Rajagopal et al., 2002; Sivakumar et al., 2007) or pupae (Bettencourt et al., 2002).

Theoretically, RNAi could target any gene associated with any tissue. Genes encoding structural PM components are an attractive target for trial of this method by "feeding" as the midgut is in close proximity to the site of dsRNA entry (Turner et al., 2006). To date, a limited number of midgut genes, including those encoding digestive enzymes (Rajagopal et al., 2002; Sivakumar et al.; 2007; Yang et al., 2010), chitinases (Khajuria et al., 2010) and chitin synthase-B (Arakane et al., 2005; 2008; Kato et al., 2006), have been silenced by RNAi. There are as yet no reports on the use of RNAi to target genes encoding peritrophins in any insect.

In the current study, I investigated the possibility of silencing midgut genes encoding peritrophins, chitin deacetylase and chitin synthase of *Mamestra configurata* identified in previous chapters. These include the genes encoding insect intestinal mucin 1 (McIIM1), insect intestinal mucin 4 (McIIM4), peritrophic matrix protein 1 (McPM1), chitin deacetylase 1 (McCDA1) and chitin synthase-B (McCHS-B). The potential for silencing each gene was first evaluated in an *in vitro* system. This revealed that *McCDA1* was silenced to the greatest extent. This demonstration was followed by *in vivo* experiments that examined the silencing of *McCDA1* in neonates and 4th instar larvae by *per os* feeding.

8.2 Material and Methods

8.2.1 Preparation of dsRNA

A region of 500, 552, 496 and 500 bp was selected and amplified for *McCDA1*, *McIIM1/4*[*], *McPM1*, and *McCHS*, respectively, using total RNA from *M. configurata* larval midgut by reverse transcription-PCR (RT-PCR). In parallel, a region of 444 bp was also amplified from the *Brassica napus* napin gene (*BnNAP*) using *B. napus* total RNA by RT-PCR. The RT-PCR products were cloned into pGEM-T Easy vector (Promega, Madison, WI, USA). The sequences of the primers used in RT-PCR are shown in Table 8.1.

[*] A region common to McIIM1 and McIIM4 was selected for synthesis of dsRNA which was denoted McIIM1/4.

Table 8.1 RT-PCR primers used for amplification of DNA templates to generate dsRNA.

Primer[1]	Sequence
BnNAP Fp	GCTCTTCCTCGTCTCGGCAACTCT
BnNAP Rp	TTCTGGAAGGGACAAACGCTAACT
McCDA1 Fp	GGACTATGGCCTTACACCCT
McCDA1 Rp	TGCATCCAGTAGTCCATTCC
McCHS-B Fp	GGCATAGTCCTCCAGATCGT
McCHS-B Rp	TCTGCTTCTTCGACGGAGGC
McIIM1/4 Fp	CCTGTCGACTTCACCATCCA
McIIM1/4 Rp	GATGCTGGAGTCAACTGGGC
McPM1 Fp	GTGCGCCGAAGGTAGACCAG
McPM1 Rp	GTTGGTCACAGCTGGGTTGG

[1]**Abbreviations:** BnNAP, *Brassica napus* napin; CDA, chitin deacetylase; CHS-B, chitin synthase-B; Fp, forward primer; IIM, insect intestinal mucin; Mc, *M. configurata*; PM; peritrophic matrix; Rp, reverse primer.

Two different clones in opposite sequence orientations were selected. Purified plasmid DNAs were linearized using restriction enzymes *Sal I* (for *McCDA1*, *McPM1* and *McCHS*) or *Pst I* (for McIIM1/4). Since *Pst I* produces 3' overhangs that are not compatible with the procedure, linearized McIIM1/4 DNA was made blunt using DNA polymerase I large (Klenow) fragment (Invitrogen, Carlsbad, CA, USA). Linearized DNA templates (3 µg) were used to generate single strand (ss) sense and antisense RNAs in a 20 µl *in vitro* transcription reaction which was conducted at 37°C for 2 h using the T7 RiboMAX Express RNAi system (Promega). ssRNAs were mixed in equal amounts, denatured at 70°C for 10 min and annealed by slow cooling at room temperature for 20 min to form the dsRNA duplex. The sample was treated with both DNase I and RNase A for 30 min at 37°C to degrade the initial linearized DNA templates and ssRNAs, respectively. The dsRNA was extracted with phenol:chloroform (1:1) and precipitated for 30 min with ice-cold 95% ethanol in the presence of 0.3 M sodium acetate at -20°C. The pellet was washed with 75% ethanol and resuspended in 25 µl of nuclease free water (Ambion, Foster City, CA, USA). The sample was run in a 1% agarose gel for verification of the dsRNA.

8.2.2 In vitro *Experiments*

Fourth instar *M. configurata* larvae were surface sterilized for 1 min in 70% ethanol, 3 min in 5% bleach and 1 min in 70% ethanol and then stored in sterile distilled water until they were dissected. Larvae were pinned on a wax plate and flooded with sterile artificial hemolymph solution (SAHS) (Palli and Locke, 1987). The cuticle was cut longitudinally and all fat body, tracheae, and Malpighian tubules were removed from the gut. The midgut was then cut longitudinally and the PM and food bolus were removed. Midguts were gently rinsed twice in SAHS and once in Grace's medium (Sigma) [with 10% fetal bovine serum (FBS) and 12.5 µg/ml gentamicin]. In total, three midguts were transferred to wells in a 24-well tissue culture plate, with each well containing 500 µl of Grace's medium with FBS and gentamicin. A total of 1.2 mg of dsRNA was applied to each well for each target (*McCDA1, McIIM1/4, McPM1* and *McCHS-B*). Three wells also containing three midguts were reserved for control treatments and did not contain dsRNA. The plate was placed on a gyrating shaker (30 rpm) at room temperature for 6 h and then the plate was transferred to a 28°C incubator. The three midguts in each well were transferred to a microcentrifuge tube containing 300 µl of TRIzol reagent (Invitrogen) after 24, 48 and 72 h. In addition, three nontreated midguts were immediately transferred into TRIzol reagent following dissection representing the 0 h control group. Samples were kept at -20°C until extraction of total RNA.

8.2.3 In vivo *Experiments*

The gene that revealed the best silencing *in vitro, McCDA1*, was used for the following *in vivo* experiments. Neonates and 4th instar larvae were used to test the success of RNAi using a droplet feeding and a leaf disk method or a combination of both.

Two sets of experiments were conducted on neonates; (1) continuous supply of dsRNA at a single dose, and (2) a single dose of dsRNA at several concentrations. The experiment based on continuous dsRNA supply was conducted using approximately 200 neonate larvae which were fed 8-10 8 µl droplets, each containing approximately 6 µg/µl of dsRNA in a 1% sucrose solution with 5% blue food coloring in nuclease free water. The larvae were allowed to feed for 1 h at room temperature. Approximately 30 larvae that were deemed to have ingested the dsRNA droplets by their blue integument when viewed under a light microscope were transferred onto *B. napus* leaf disks. The disks had been previously covered on each side with 250 µg of dsRNA in

50 µl 0.1% Triton-100 in nuclease free water. Leaf disks with dsRNA were replaced every 24 h with fresh disks. Control larvae were fed nuclease free water containing only blue food coloring in sucrose solution and were transferred onto *B. napus* leaf disks with 50 µl 0.1% Triton-100 on each side without dsRNA. The second experiment on neonates was conducted using only the droplet feeding method as explained above, but with dsRNA concentrations of 0.375, 0.75, 1.5, 3, 6 and 12 µg/µl. Larvae were then supplied with only leaf disks to continue their feeding. In these experiments, the effect of a dsRNA specific to an unrelated gene (*B. napus BnNAP*) on the expression of *McCDA1* was also examined at the same concentrations of dsRNA specific to *McCDA1*. Therefore, control larvae were fed nuclease free water containing blue food coloring in sucrose solution with or without dsRNA specific to *BnNAP* and were then supplied with only leaf disks to continue their feeding. Larvae were collected at 24 and 48 h, transferred into microcentrifuge tubes containing 300 µl TRIzol reagent, and kept at -20°C until extraction of total RNA.

The experiment on 4th instar larvae was conducted using the leaf disk method as explained above with slight differences. Individual 4th instar larvae were fed a *B. napus* leaf disk with 125 µg of dsRNA in 25 µl 0.1% Triton-100 in nuclease free water; however, leaf disks were changed every 12 h. Control larvae were fed leaf disks with 25 µl 0.1% Triton-100 in nuclease free water without dsRNA which were also changed every 12 h. Midguts were dissected from individual larvae after 24 and 36 h and transferred into microcentrifuge tubes containing 300 µl TRIzol reagent and kept at -20°C until extraction of total RNA. In addition, the dsRNA construct for *McCHS-B* was also tested *in vivo* on 4th instar larvae as described above for *McCDA1*.

8.2.4 RNA Extraction and Analysis of Gene Expression

Expression of target genes was examined using RT-PCR with total RNA extracted from entire neonates or individual midguts dissected from 4th instar larvae as described in Chapter 2. Primers used in RT-PCR for expression of *McCDA1*, *McIIM1*, *McIIM4*, *McPM1* and *McTUB* are shown in Table 5.2 and Table 3.2 for *McCHS-B*.

8.3 Results

8.3.1 In vitro *Experiments*

RT-PCR analysis of midguts treated with dsRNA specific to *McCDA1* showed that *McCDA1* mRNA levels declined at 24 h post-treatment (hpt), and that no transcripts were detectable by 48 and 72 hpt (Figure 8.1). The level of *McCDA1* mRNA in untreated midguts declined only gradually with time, as was the case for *McTUB* transcripts, indicating that the silencing was specific to *McCDA1* (Figure 8.1). *McPM1* mRNA was detected in midguts treated with dsRNA specific to *McPM1* at 24 and 48 hpt; however, no *McPM1* mRNA was detectable by 72 hpt (Figure 8.1). As before, the levels of *McTUB* mRNA in the control and treated midguts declined only slightly with time, as did *McPM1* mRNA levels in the untreated midguts, indicated that silencing was specific to *McPM1*. The reduced transcript levels in the control samples at 72 h may reflect a decline in the health of the cells or loss of midgut gene expression patterns. The decrease in the expression of all genes was more obvious by 96 h (data not shown); therefore, data obtained in these *in vitro* experiments are most informative after 24-72 hpt.

Transcripts derived from *McIIM1* and *McIIM4* were detected throughout the 72 h experiment in midguts treated with a dsRNA designed to silence both genes. However, the amount of transcript, especially that of *McIIM1,* was lower at 72 hpt in the treated midguts compared to the untreated control midguts (Figure 8.1), indicating a weak silencing effect. *McCHS-B* mRNA levels were unaffected, indicating that the *McCHS-B* was not silenced by treatment with the dsRNA construct used in this experiment (Figure 8.1).

8.3.2 In vivo *Experiments*

RT-PCR analysis using total RNA from neonates continuously fed *McCDA1* dsRNA revealed that *McCDA1* mRNA was undetectable after 24 and 48 hpt (Figure 8.2). In contrast, *McTUB* mRNA levels remained fairly constant in these larvae and *McCDA1*mRNA levels were unaffected in untreated neonates, indicating that the silencing was specific to *McCDA1* (Figure 8.2). RT-PCR analysis using midguts from 4th instar larvae continuously fed *McCDA1* dsRNA showed that they were less affected by the treatments and a decline in transcript levels was only detected at 36 hpt (Figure 8.2). However, continuous feeding of 4th instar *M. configurata* larvae with 125 µg of *McCHS-B* dsRNA did not result in any silencing at 24 or 48 hpt (data not shown).

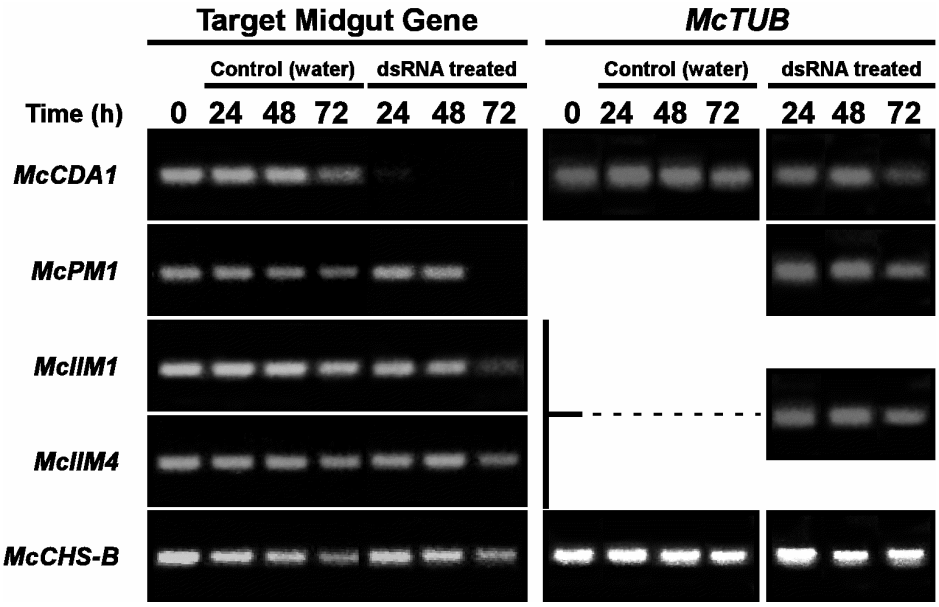


Figure 8.1 RT-PCR analysis to detect mRNA from *Mamestra configurata* midgut genes (*McCDA1, McPM1, McIIM1, McIIM4* and *McCHS-B*) targeted by dsRNA *in vitro*. A total of 1.2 mg of dsRNA specific to each gene was applied to three midguts in 500 µl of Grace's medium containing 10% fetal bovine serum and 12.5µg/ml gentamicin. Nuclease free water was used in the control treatments instead of dsRNA. Amplification of the *M. configurata* tubulin (*McTUB)* gene was used as a control in the RT-PCR to ensure equivalent RNA template in samples.

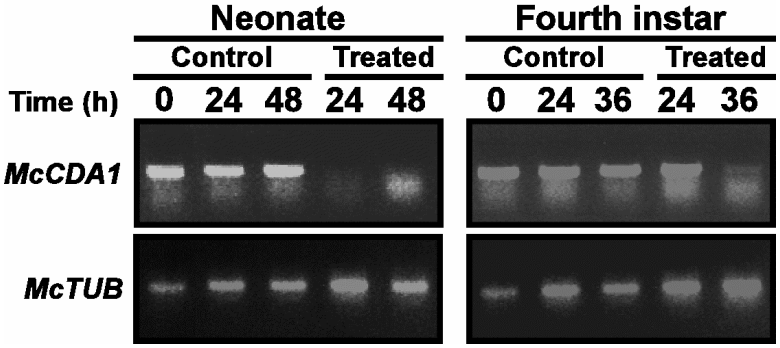


Figure 8.2 RT-PCR analysis to detect the expression of *Mamestra configurata McCDA1* from neonates or 4th instar larvae continuously fed with dsRNA specific to *McCDA1*. Neonates were fed droplets containing approximately 6 µg/µl of dsRNA in a 1% sucrose solution with 5% blue food coloring in nuclease free water for 1 h and were supplied continuously with fresh *B. napus* leaf disks, covered on each side with 250 µg of dsRNA in 50 µl 0.1% Triton-100 in nuclease free water without dsRNA. Fourth instar larvae were supplied continuously with fresh *B. napus* leaf disks with 125 µg of dsRNA in 25 µl 0.1% Triton-100 in nuclease free water. Control larvae were fed nuclease free water containing only blue food coloring in sucrose solution and were transferred onto leaf disks with 25 or 50 µl 0.1% Triton-100 on each side. Total RNA was extracted from 30 neonate larvae at 24 and 48 h or from midguts dissected from 4th instar larvae at 24 and 36 h. Amplification of the *M. configurata* tubulin (*McTUB*) gene was used as a control in the RT-PCR to ensure equivalent RNA template in samples.

Groups of neonate larvae were fed once with droplets with dsRNA at 0.375, 0.75, 1.5, 3, 6 and 12 µg/µl specific to *McCDA1* or *B. napus napin* (*BnNAP*). The 12 µg/µl concentration of either dsRNA appeared to be toxic to the larvae as they did not move, eat or react to their surroundings for at least 24 h after feeding. At lower concentrations, *McCDA1* mRNA was undetectable after 24 h in neonates fed 6 µg/µl of dsRNA, with only a slight amount of mRNA detectable at 48 hpt., mRNA levels were greatly reduced in neonates fed 1.5 or 3 µg/µl at 48 hpt (Figure 8.3). The lowest concentrations of dsRNA (0.375 and 0.75 µg/µl) did not have any silencing effect on *McCDA1* (data not shown). *McCDA1* mRNA levels were unaffected in control larvae fed nuclease-free water without dsRNA or dsRNA specific for *BnNAP* indicating that the silencing was due to *McCDA1* dsRNA uptake. *McTUB* mRNA levels remains constant in all treatments.

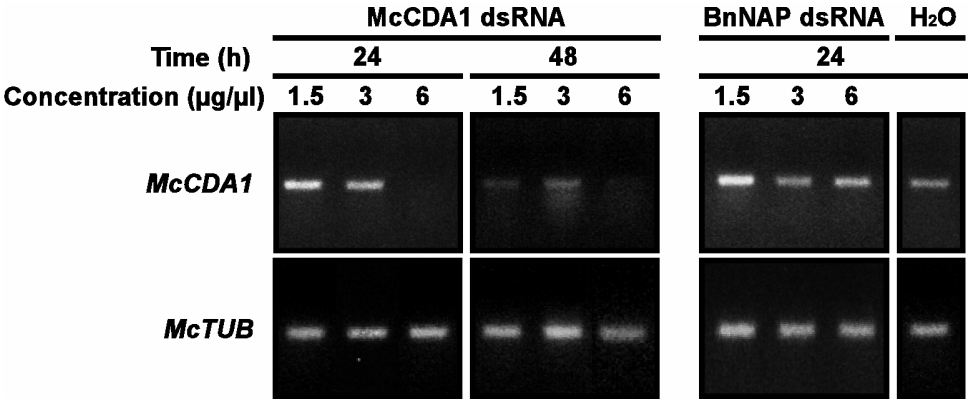


Figure 8.3 RT-PCR analysis to detect the expression of *Mamestra configurata McCDA1* from neonates fed a single dose of dsRNA specific to *McCDA1* at several concentrations. Neonates were fed droplets containing approximately 1.5, 3, and 6 µg/µl of dsRNA in a 1% sucrose solution with 5% blue food coloring in nuclease free water for 1 h. Control larvae were fed nuclease free water containing blue food coloring in sucrose solution with or without dsRNA specific to the *B. napus napin* (*BnNAP*) gene. All larvae were supplied with leaf disks following droplet feeding in order to continue on their feeding. Total RNA was extracted from 30 neonate larvae at 24 and 48 h. Amplification of the *M. configurata* tubulin (*McTUB*) gene was used as a control in the RT-PCR to ensure equivalent RNA template in samples.

8.4 Discussion

In vitro studies revealed an acute silencing of *McCDA1* within 24 hpt that persisted up to 72 hpt. Strong silencing of *McPM1* was also detected by 72 hpt, while silencing of *McIIM1* and *McIIM4* was weak. In contrast to the *McCDA1* and peritrophin genes, no silencing was detected for *McCHS-B*. The reason for this variation in the silencing effect between these genes is not

clear as the same dosages and dsRNAs of similar sizes were used under the same conditions for all target genes. Another variable in RNAi experiments is the nucleotide content of the gene regions selected for dsRNA synthesis. It has been commonly observed that high guanine (G) and cytosine (C) content negatively correlates with RNAi activity and it is recommended that sequences with low GC content (30-55%) be used (Chan et al., 2009). The GC content of the selected regions for dsRNA synthesis was 53%, 52.4%, 46.6%, 53% for *McCDA1*, *McIIM1/4*, *McPM1*, *McCHS*, respectively, suggesting that GC content is not likely to be the reason for the variation detected in the current study. mRNA abundance of the target genes was also reported as a factor affecting RNAi success (Tomoyasu and Denell, 2004). Genes expressed at lower levels may be targeted more efficiently than the genes expressed in higher levels by the same doses of dsRNA; thus, it may be necessary to optimize the amount of dsRNA for each gene (Tomoyasu and Denell, 2004). However, the amount used in the current study was quite high compared to other studies, suggesting this is not be the reason for the variation in RNAi efficiency.

Per os feeding of dsRNA to neonates or 4th instar larvae showed silencing of *McCDA1* starting at 24 or 36 hpt, respectively, in accordance with the results of the *in vitro* experiment. This is one of the most successful demonstrations of silencing of a lepidopteran gene by RNAi (Terenius et al., 2011); however, it must be recognised that larvae were fed continuously with a high concentration (6 µg/µl initial uptake followed by 250 µg/leaf disk for neonates and 125µg/leaf disk for 4th instar larva) of dsRNA. In contrast, RNAi studies from coleopterans (Brown et al., 1999; Bucher et al., 2002; Tomoyasu and Denell, 2004; Zhu et al., 2008b) and dipterans (Misquitta and Paterson 1999; Eaton et al., 2002; Kato et al., 2006) reported efficient gene silencing after a single injection of 5 pg-500 ng of dsRNA. When the larvae were fed once, dsRNA at a concentration of 6 µg/µl gave efficient gene silencing while higher concentrations (12 µg/µl) appeared to be toxic to the larvae. Larger amounts of dsRNA appear to be necessary for lepidopterans as Rajagopal et al (2002) and Sivakumar et al. (2007) showed silencing of a midgut aminopeptidase gene using 4 and 6 µg of dsRNA/larva for 5th instar *Spodoptera litura* and *Helicoverpa armigera* larvae, respectively. Turner et al. (2006) and Yang et al. (2010) reported approximately 25-50% reductions in the expression of gut carboxylesterase and aminopeptidase genes within 2 days after feeding dsRNA to 3rd instar *Epiphyas postvittana* (1 µg) and *Diatraea saccharalis* (250 ng) larvae, respectively. Lower concentrations (≤ 3µg/µl) did not cause detectable silencing as measured by RT-PCR in the current study, supporting the

statement that RNAi requirements vary among species, instars, and even individuals within a species (Geldhof et al., 2006).

The current study is the first demonstration of silencing of any gene encoding a PM protein, in this case McCDA1 and McPM1, *in vitro* and/or *in vivo*. However, there have been several demonstrations of silencing midgut genes encoding chitinase and chitin synthase, but only one of these studies is from a lepidopteran. Continuous feeding of neonate *Ostrinia nubilalis* larvae with 10 µg of dsRNA (a total of 30 µg dsRNA/larva) reduced midgut chitinase expression by over 60% (Khajuria et al., 2010). Other studies reported silencing of *CHS-B* upon injection of 200 ng dsRNA/insect in larval (Arakane et al., 2005) or adult stages (Arakane et al., 2008) of *Tribolium castaneum*, or 500 ng of dsRNA/insect in *Aedes aegypti* adults (Kato et al., 2006). I also tried to silence *McCHS-B in vivo* even though *in vitro* experiments were unsuccessful. Feeding 4th instar *M. configurata* larvae with dsRNA did not result in any *McCHS-B* silencing at 24 or 48 hpt, suggesting that testing in an *in vitro* system may be a good screen for selecting dsRNA constructs before proceeding to *in vivo* studies.

One of the factors limiting the utility of RNAi in lepidopterans is the lack of persistent silencing. Injection of dsRNAs into larvae (Tomoyasu and Denell, 2004; Arakane et al., 2005) or adult (Arakane et al., 2008) coleopterans caused an excellent silencing effect that was retained even after molting. Only a few studies reported that gene silencing was retained after molting and even transmitted to the subsequent generation in lepidopterans, for example, silencing of a *S. litura* midgut aminopeptidase gene (Rajagopal et al., 2002) and *Hyalophora cecropia* hemolin gene (Bettencourt et al., 2002). However, the mechanisms behind the systemic and persistent aspects of larval RNAi are still unknown.

The midgut of most lepidopterans is a hostile environment for RNA because of its alkaline pH and the presence of RNase activity (Terra and Ferreira, 1994). Interestingly, this does not seem to affect the stability of ingested dsRNA as it must remain intact in the insect midgut lumen to enter the midgut cells and initiate silencing (Turner et al., 2006). Expression of a dsRNase in the *M. configurata* midgut and its association with the PM (Chapter 4) is noteworthy as it may be interesting to investigate whether this enzyme is a factor in the requirement to feed much higher levels of dsRNA to lepidopterans to initiate RNAi. Variation in the midgut environment between species may also determine if conditions are conducive for

RNAi to be initiated by dsRNA feeding and not all lepidopteran species may be suitable for these types of experiments.

In conclusion, the current study revealed an acute and rapid silencing for *McCDA1* and a slower rate of silencing for *McPM1* upon *in vitro* treatment of the midgut with dsRNAs that are approximately 500 bp long. *McIIM1* and *McIIM4* were slightly silenced by 72 hpt, while no silencing was detected for *McCHS-B*. Continuous feeding of neonates and 4th instar larvae with dsRNA resulted in silencing of *McCDA1*. Feeding of a single dose to neonates also resulted in silencing of *McCDA1*. RNAi in insects has the potential to be part of new insect control strategies (Asokan, 2008) and the use of transgenic plants expressing dsRNA for target genes is very promising (Baum et al., 2007). *McCDA1* has potential in this regard because it encodes an active CDA (Chapter 5) and RNAi could successfully target its expression.

9. GENERAL DISCUSSION, CONCLUSIONS, AND FUTURE DIRECTIONS

Although the presence of a peritrophic matrix (PM) in insects has been known for over two centuries, detailed studies on PM components and their contribution to PM function began only in late 1990s. Much progress has been achieved in the last 5 years with the advent of genomic and proteomic sciences. The current study has enhanced the understanding of lepidopteran PM morphology, formation, biochemistry, and function. In addition, the two studies related to the baculoviral metalloprotease, enhancin, and the use of RNA interference (RNAi) to silence midgut gene expression have shown the potential for the development of novel insect control strategies.

In *Mamestra configurata*, a multilayered PM is continuously present in feeding and non-feeding (molting and starving) stages, with the exception of the Phase I molting stage. This is in accordance with the constant expression of *M. configurata* chitin synthase-B, *McCHS-B*, during larval development. In contrast, in female mosquitoes, chitin is secreted in response to feeding, and the PM is visible almost immediately (Freyvogel and Jaquet, 1965), corresponding to the upregulation of *CHS-B* (Ibrahim et al., 2000). At the microscopic level, PMs from feeding *M. configurata* larvae have putative chitin bundles, observable as striations running perpendicular to the long axis of the PM in feeding larvae. The presence of a multilayered PM in combination with perpendicular chitin bundles suggests that the PM periodically delaminates from the midgut epithelium in the anterior midgut, which may correspond to occasional periods when feeding ceases (resting) followed by pulses of PM synthesis. Indeed, examination of PMs from *Trichoplusia ni* (Harper and Granados, 1999), *Ostrinia nubilalis* (Harper and Hopkins, 1997) and *Manduca sexta* (Hopkins and Harper, 2001) using gold labelled wheat germ agglutinin revealed that the anterior midgut is the site of the chitin secretion. In addition, the *CHS-B* gene is expressed at higher levels in the anterior midgut (Zimoch and Merzendorfer 2002; Arakane et al., 2004).

Broehan et al. (2007) proposed that chitin synthesis is controlled by a cascade involving CHS-B and a putative serine protease activator, chymotrypsin like protease 1 (CTLP1) in *Manduca sexta*. Both *McCHS-B* and the *CTLP1* ortholog, *M. configurata* serine protease 33 gene (*McSP33*), are expressed at constant levels during larval development, suggesting an even more sophisticated mechanism is at play. The *M. configurata* midgut expresses a splice variant of the *M. configurata* serine protease inhibitor 1 (McSerpin1) gene, McSerpin1A (Chamankhah et al.,

2003), which could inhibit the activity of the McSP33. *McSerpin1* is downregulated in molting Phase II (Chamankhah et al., 2003) during which a new PM is formed. Similarly, *McSerpin1* is upregulated in response to starvation (Chamankhah et al., 2003) and this would prevent CHS-B activation by the McSP33 and limit chitin synthesis.

In molting Phase I, genes encoding the chitinolytic enzymes *M. configurata* chitinase (McCHI) and β-*N*-acetylglucosaminidase (McNAG) are dramatically upregulated and NAG activity peaks, coincident with degradation of the old PM. In molting Phase II, a new PM with an inner opaque layer surrounded by an outer translucent layer appears. The inner layer is the initial PM synthesized following molting Phase I and is more mature. It may also have a higher chitin content as it reacts strongly with the optical brightener fluorescent brightener 28 (FB28). The outer layer facing the midgut epithelial cells is translucent and likely to have a lower chitin content, as evidenced by reduced binding of FB28. Therefore, it is a product of more recent secretion and represents an immature PM.

Although dramatic upregulation of *McCHI* and *McNAG* genes in molting Phase I along with the increased NAG activity suggests a transcriptional based control of chitin degradation, the dramatic decrease in NAG activity despite its high level of transcription in molting Phase II is contradictory to this hypothesis. Thus, chitinolytic activity by McNAG may not be controlled only at the level of transcription, but possibly also at the level of protein activation by serine proteases (Koga et al., 1989). This is also required for the protection of the newly formed PM in molting Phase II.

Proteomic analysis identified 82 proteins that are strongly associated with the *M. configurata* PM; these could be classified as peritrophins, enzymes [chitin deacetylases (CDAs) and digestive enzymes] and other proteins. dsRNase, astacin, pantetheinase, response to pathogen (REPAT) protein, midgut 176 (MG176), C type lectin (CLECT) as well as three PM proteins without orthologs in other species, represent the first report of their presence in insect PM. Expression of the majority of the genes encoding these proteins is specific to the midgut, with the exceptions of *M. configurata* β-1, 3-glucanase (*Mcβ1, 3GLU*), pantetheinase (*McPAN*), *McCLECT, McSerpin1A, 363-100-1* and *357-80-1*, which may also be expressed in the foregut, midgut, hindgut, Malpighian tubules, tracheae, integument and fat body.

Being structural components of the PM, peritrophins represent one of the most important types of PM proteins (Tellam et al., 1999). Peritrophins can be classified into four classes

(simple, binary, complex and repetitive) according to their structural organization and phylogenetic analysis of their peritrophin A domains (PADs). The widespread presence of each class in lepidopterans suggests a conserved evolutionary development for each class of peritrophins. In addition, all lepidopteran peritrophins reported to date have a PAD, which is a Type 2 chitin binding domain (CBD). However, one of the nonmucin proteins identified here, *M. configurata* chitin binding domain 3 protein (McCBD3P), has a Type 3 CBD. This is the first report of such a domain in lepidopteran peritrophins. Recently, a protein with this type of CBD was also reported to be associated with the *Anopheles gambiae* adult PM (Dinglasan et al., 2009), suggesting they may be a common feature in insect PMs.

One of the more interesting findings on peritrophins in this study is the variation in their sites(s) of expression. For example, genes encoding *M. configurata* peritrophic matrix 1 (McPM1) and *M. configurata* protein 1 with peritrophin A domain (McPPAD1) are expressed in foregut, hindgut, Malpighian tubules, tracheae, integument and fat body, whereas the gene encoding McCBD3P is expressed in tracheae and Malpighian tubules in addition to midgut. These findings clearly indicated that expression of some peritrophin genes is not restricted to tissues synthesizing PM, although expression of *M. configurata* insect intestinal mucin (McIIM) genes is restricted to midgut. However, both IIM and nonmucin peritrophin genes are expressed continuously, indicating their expression is independent of larval feeding status, developmental stage, or PM formation. In addition, lepidopteran peritrophins are synthesized in the anterior midgut (Bolognesi et al., 2001; Campbell et al., 2008) before chitin fibers form the PM network (Ryerse et al., 1992; Harper and Hopkins, 1997; Harper and Granados, 1999). These peritrophins may bind to pre-existing immature PM in the anterior midgut during the process of PM maturation (Tellam et al., 2000), suggesting the inner opaque PM detected in molting Phase I may already contain the peritrophins, thus explaining its amorphous structure. This association may stabilize the PM by interlocking the chitin fibrils quickly and protect the PM from potential hazardous factors.

CDAs were first discovered in other chitinous tissues, such as the integument (Neville, 1975) and tracheae (Wang et al., 1996), and were later found in lepidopteran PM (Guo et al., 2005). The current study is the first report of CDA activity in the insect midgut and the first demonstration that a recombinant insect CDA is active, suggesting that the PM may also contain chitosan. Deacetylation of chitin increases its solubility and decreases the density of chitin fibrils

in vitro (Cho et al., 2000), which may result in alterations in PM integrity and permeability (Jakubowska et al., 2010). This may be important for the flexibility of the PM within the midgut. Continuous expression of CDA genes (current study) along the entire midgut (Campbell et al., 2008) suggests that deacetylation may occur wherever and whenever it is needed.

Digestive enzymes such as insect intestinal lipases (IILs) and serine proteases are also associated with the PM. In *M. configurata*, the various IIL genes differ in their expression during larval development. However, serine protease genes are expressed continuously and serine protease activity is present in the midgut of feeding and nonfeeding (starving and molting) larvae. The association of digestive enzymes with the PM may be attributed to simple entrapment as they cross the PM to the endoperitrophic space. However, the fact that they are so strongly associated with PM suggests that they are fixed and contribute to digestion of polypeptides or fats moving through the PM towards the epithelium. Many of the serine proteases and IILs appear to be inactive as they lack critical catalytic amino acids in the enzyme active site. In nature, inhibitors of protease (Dunaevsky et al., 2005) or lipases (Chapman, 1987) are commonly found in plants. The inactive serine proteases or lipases may bind to inhibitors from plant tissue and reduce the interaction of these inhibitors with active enzymes (Mazumdar-Leighton et al., 2000; Christeller et al., 2010). Inactive IILs may also bind to free fatty acids and help to carry them to the midgut epithelium for absorption (Christeller et al., 2010).

In conclusion, findings from the current study together with those previously published (Ryerse et al., 1992; Harper and Hopkins, 1997; Harper and Granados, 1999; Bolognesi et al., 2001; Campbell et al., 2008) suggest that the lepidopteran PM is intermediate between Types I and II as chitin and peritrophins are synthesized by the anterior midgut while CDAs are produced across the midgut. Formation of the lepidopteran PM occurs in two steps during molting Phase II: (1) periodic formation of immature or primal PMs by delamination from the anterior midgut epithelium by pulses, (2) maturation of the PM by deacetylation as it passes towards the posterior midgut (Figure 9.1).

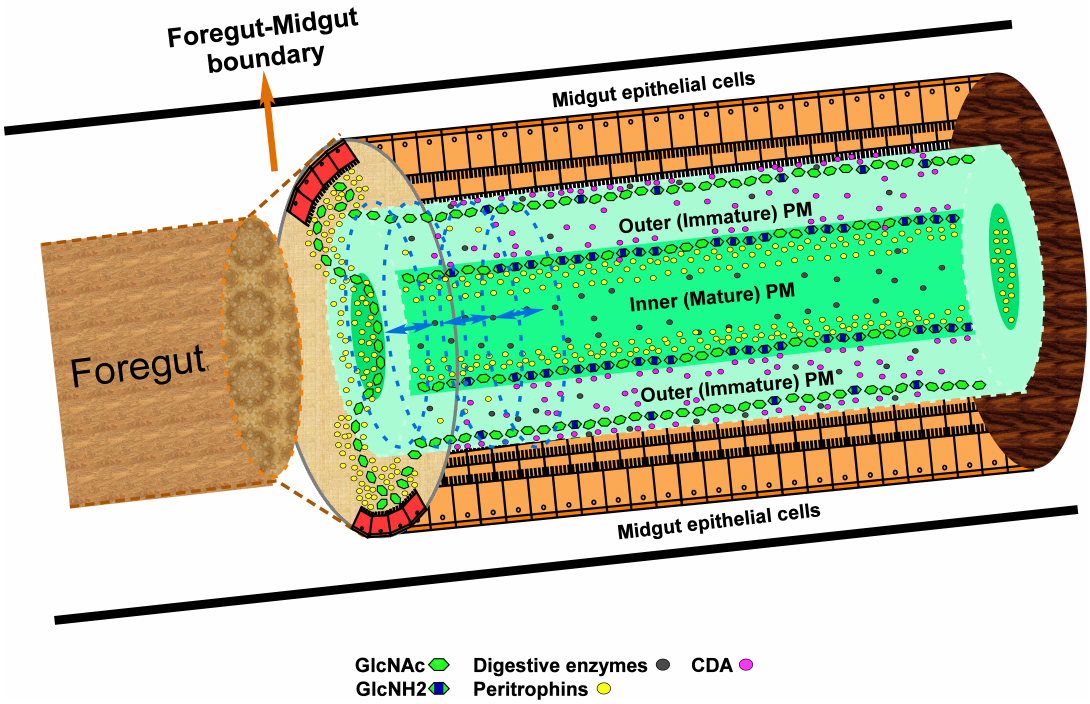


Figure 9.1 Model outlining the processes involved in formation of the lepidopteran peritrophic matrix (PM) in molting Phase II. The sites of *N*-acetyl-D-glucosamine (GlcNAc), D-glucosamine ($GlcNH_2$), chitin deacetylase (CDA), peritrophin and digestive enzyme synthesis and incorporation are indicated. Blue arrow indicates the periodic pulses of formation originating from the anterior midgut epithelium.

This study identified several new PM proteins thus warranting revision of the molecular model of PM (Figure 9.2). As in earlier models, repetitive peritrophins, such as McPM1 (Shi et al., 2004), *T. ni* chitin binding protein 1 (CBP1), and chitin binding protein 2 (CBP2) (Wang et al., 2004a) are proposed to hold the chitin fibrils and bundles rigid. Because CBDs are held together by several disulphide bridges, access to protease cleavage sites is restricted. Indeed, treatment with dithiothreitol, a disulphide bond reducing agent, led to the degradation of *T. ni* IIM by endogenous digestive proteases (Wang and Granados, 1997b). Additional disulphide bridges formed by the cysteines that lie outside CBDs may contribute to protein multimerization similar to the multimerization of vertebrate mucins by intermolecular disulphide bridges (Perez-Vilar and Hill, 1999). The second important feature of the model is related to *O*-glycosylation of IIMs. Their high degree of glycosylation serves to protect the PM from physical and proteolytic damage (Wang and Granados, 1997a). The third important feature is related to the role of *N*-glycosylation of asparagines that are common in the spacer regions. At the core of the *N*-linked

glycan, two GlcNAc residues are linked by a β-(1, 4) bond as in chitin; suggesting the di-*N*-acetylglucosamine in the core glycan could be recognized by a CBD of another peritrophin, resulting in formation of protein complexes (Shi et al., 2004). Spacer elements may contribute to the PM integrity in several other ways. The presence of glycines and prolines in the spacers of IIMs [e.g., *Epiphyas postvittana* IIM1 (Simpson et al., 2007), *Helicoverpa armigera* IIM1 (Campbell et al., 2008) and *Plutella xylostella* IIM (Sarauer et al., 2003)] may facilitate the formation of tightly coiled structures that bring the CBDs into close proximity to one another (Shi et al., 2004). In addition, the preponderance of negatively charged amino acids in these spacer regions may prevent the interaction of serine proteases with PM proteins. It is also possible that these negatively charged groups may help to adsorb IIMs to the PM by interacting with the amino groups of chitosan, further increasing structural integrity. The fourth feature of the model is related to the simple peritrophins such as McPPAD1 or McCBD3P. These small proteins may fill the gaps remaining after assembly of microfibrils. Peritrophin-15, a simple peritrophin with a single CBD identified from *Chrysomya bezziana* and *Lucilia cuprina*, binds to chitin at a low stoichiometry, suggesting that such proteins cap the end of individual chitin chains and protect the fibrils from exochitinases (Wijffels et al., 2001). Similarly, McPPAD1 or McCBP3 may protect the ends of the chitin chains from degradation.

The current study revealed that *M. configurata* IIM4 (McIIM4) was susceptible to degradation by Mamestra configurata nucleopolyhedrovirus-A (MacoNPV-A) challenge, as the first evidence of IIM degradation by alphabaculovirus enhancin; though IIM degradation by betabaculovirus enhancins was previously shown (Wang and Granados, 1997a). The fact that alphabaculovirus enhancin is localized to occlusion derived virions (ODVs) (Slavicek and Popham, 2005) and betabaculovirus enhancin to occlusion bodies (OBs) (Roelvink et al., 1995) suggests that their impact on PM may differ. Alphabaculovirus enhancin action on IIMs is likely localized to the site of interaction of ODVs with the PM, in contrast to the more widespread degradation caused by the release of betabaculovirus enhancin after solubilization of the OB in the midgut. *M. configurata* IIM2 (McIIM2), a binary peritrophin, is unaffected by baculoviral challenge and such resistance of an IIM has not been reported previously. The susceptible IIMs, such as *M. configurata* IIM1 (Shi et al., 2004), McIIM4 (Chapter 7) and TnIIM (Wang and Granados, 1997a) are all members of the complex peritrophins which have lower relative levels of *O*-glycosylation, in comparison to McIIM2. This suggests that a major factor affecting IIM

degradation is glycosylation. Interestingly, binary peritrophins are found in all lepidopterans examined and have a high *O*-glycosylation potential, suggesting that they have a conserved and critical function; this may include protection against pathogen invasion.

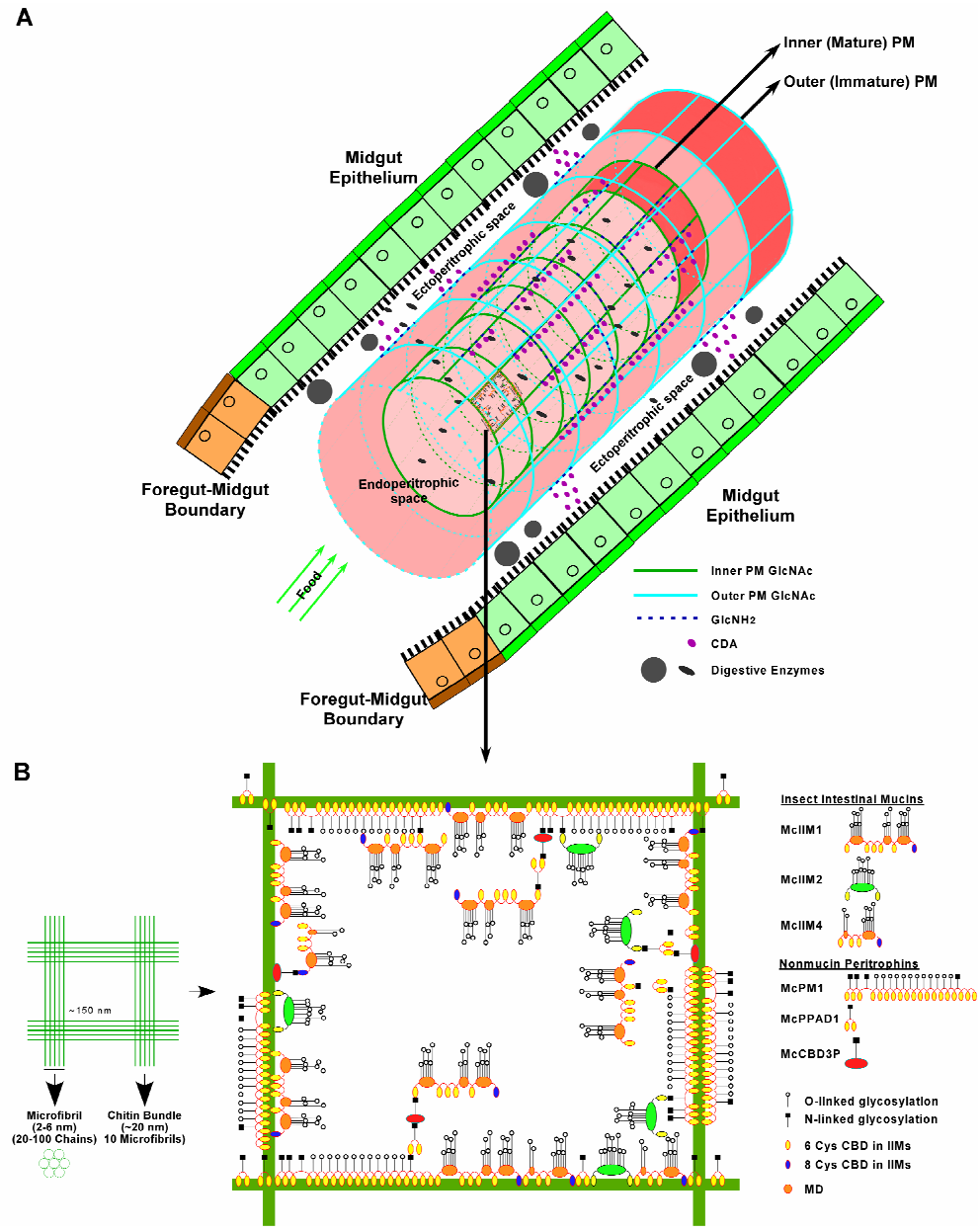


Figure 9.2 Model of the lepidopteran peritrophix matrix (PM) formation (A) and its molecular architecture (B). Chitin microfibrils consisting of 20 to 400 chitin chains assemble into bundles to form a scaffold. Insect intestinal mucins are portrayed with the *O*-linked glycans projecting into the interstitial space. CBDs are also shown interacting with the β-(1,4)-*N*-acetyl-d-glucosamine dimer at the core of *N*-linked glycans to generate additional intermolecular interactions. The model in panel B is adapted from Shi et al. (2004).

The current study is also the first demonstration of silencing of any gene encoding a PM protein by RNAi, in this case *M. configurata* CDA1 (McCDA1) and McPM1. Thus, both *in vitro* and *per os* feeding experiments revealed *McCDA1* silencing starting at 24 or 36 hpt, as one of the most successful demonstrations of RNAi in lepidopterans (Terenius et al., 2011). In addition, RNAi has been shown to have great potential in insect control by development of transgenic plants expressing dsRNA (Baum et al., 2007). Therefore, silencing of *McCDA1* through transgenic plants may have great potential as a novel control strategy for *M. configurata*.

My research has contributed to a detailed level of understanding of the formation, structure, maintenance, and roles of the insect PM. The efforts not only of this but other entomological research projects have an ultimate aim to develop or at least to contribute to potential novel insect control strategies. This goal is not necessary only to diminish the extensive crop loses by insect pests but also necessary for development of environment friendly control tools. Thus, the laws of the future will not allow use of most of the toxic chemical insecticides that are currently used extensively. The PM has great potential as a target because it has essential roles in insect digestive physiology and is indirectly in contact with the external environment, which allows a *per os* application approach. Several strategies were tried during the course of my research but only two of them showed potential, the use of baculovirus enhancin to target mucins and RNAi to target CDAs. Transgenic plants encoding enhancin or dsRNA specific to *McCDA1* should be evaluated in terms of their effects on *M. configurata*. In addition, optimization of an RNAi method by *per os* feeding for lepidopterans may also be an important contribution to this field. Although this study has elucidated important features of major PM components such as peritrophins, CDAs, and enzymes involved in digestion or chitin metabolism, the role of some of the components identified by proteomics remained unresolved. In particular, presence of astacin, dsRNase, polycalin, REPAT, CLECT, and orthologs of MG176, Lsti99 and Lsti201 proteins and three novel *M. configurata* proteins (357-94-1, 530-247-1, and 357-80-1) that had no orthologs in GenBank are proteins of potential interest. Investigating the role(s) and structure of these components may allow development of further potential targets in insect control.

REFERENCES

Abraham, E.G., Jacobs-Lorena, M., 2004. Mosquito midgut barriers to malaria parasite development. Insect Biochem. Mol. Biol. 34, 667-671.

Adang, M.J., Spence, K.D., 1981. Surface morphology of peritrophic membrane formation in the cabbage looper, *Trichoplusia ni*. Cell Tissue Res. 218, 141-147.

Adang, M.J., Spence, K.D., 1982. Biochemical comparisons of the peritrophic membranes of the lepidopterans *Orgyia pseudotsugata* and *Manduca sexta*. Comp. Biochem. Physiol. 73B, 645-649.

Ahmad, T., Rajagopal, R., Bhatnagar, R.K., 2003. Molecular characterization of chitinase from polyphagous pest *Helicoverpa armigera*. Biochem. Biophys. Res. Commun. 310, 188-195.

Alger, N.E., Cabrera, E.J., 1972. An increase in death rate of *Anopheles stephensi* fed on rabbits immunized with mosquito antigen. J. Econ. Entomol. 65, 165-168.

Amdam, G.V., Simões, Z.L.P., Guidugli, K.R., Norberg, K., Omholt, S.W., 2003. Disruption of vitellogenin gene function in adult honeybees by intraabdominal injection of double-stranded RNA. BMC Biotechnol. 3, 1.

Angelucci, C., Barrett-Wilt, G.A., Hunt, D.F., Akhurst, R.J., East, P.D., Gordon, K.H.J., Campbell, P.M., 2008. Diversity of aminopeptidases, derived from four lepidopteran gene duplications, and polycalins expressed in the midgut of *Helicoverpa armigera*: Identification of proteins binding the d-endotoxin, Cry1Ac of *Bacillus thuringiensis*. Insect Biochem. Mol. Biol. 38, 685-696.

Arakane, Y., Muthukrishnan, S., 2010. Insect chitinase and chitinase-like proteins. Cell. Mol. Life Sci. 67, 201-216.

Arakane, Y., Dixit, R., Begum, K., Park, Y., Specht, C.A., Merzendorfer, H., Kramer, K.J., Muthukrishnan, S., Beeman, R.W., 2009. Analysis of functions of the chitin deacetylase gene family in *Tribolium castaneum*. Insect Biochem. Mol. Biol. 39, 355-365.

Arakane, Y., Hogenkamp, D.G., Zhu, Y.C., Kramer, K.J., Specht, C.A., Beeman, R.W., Kanost, M.R., Muthukrishnan, S., 2004. Characterization of two chitin synthase genes of the red flour beetle, *Tribolium castaneum*, and alternate exon usage in one of the genes during development. Insect Biochem. Mol. Biol. 34, 291-304.

Arakane, Y., Muthukrishnan, S., Kramer, K.J., Specht, C.A., Tomoyasu, Y., Lorenzen, M.D., Kanost, M.R., Beeman, R.W., 2005. The *Tribolium* chitin synthase genes *TcCHS1* and

TcCHS2 are specialized for synthesis of epidermal cuticle and midgut peritrophic matrix, respectively. Insect Mol. Biol. 14, 453-463.

Arakane, Y., Specht, C.A., Kramer, K.J., Muthukrishnan, S., Beeman, R.W., 2008. Chitin synthases are required for survival, fecundity and egg hatch in the red flour beetle, *Tribolium castaneum*. Insect Biochem. Mol. Biol. 38, 959-962.

Arakane, Y., Zhu, Q., Matsumiya, M., Muthukrishnan, S., Kramer, K.J., 2003. Properties of catalytic, linker and insoluble substrate binding domains of insect chitinase. Insect Biochem. Mol. Biol. 33, 631-648.

Arimatsu, Y., Furuno, T., Sugimura, Y., Togoh, M., Ishihara, R., Tokizane, M., Kotani, E., Hayashi, Y., Furusawa, T., 2007a. Purification and properties of double-stranded RNA-degrading nuclease, dsRNase, from the digestive juice of the silkworm, *Bombyx mori*. J. Insect Biotechnol. Sericol. 76, 57-62.

Arimatsu, Y., Kotani, E., Sugimura, Y., Furusawa, T., 2007b. Molecular characterization of a cDNA encoding extracellular dsRNase and its expression in the silworm, *Bombyx mori*. Insect Biochem. Mol. Biol. 37, 176-183.

Asokan, R., 2008. RNA interference (RNAi): A novel strategy in insect pest management. Curr. Sci. 94, 1119.

Aubertot, M., 1932. Les sacs peritrophiques des larves d'*Aeshna*; leur evacuation periodique. C. R. Soc. Biol. 111, 746-748.

Baines, D.M., 1978. Observations on the peritrophic membrane of *Locusta migratoria migratoroides* (R. and F.) nymphs. Acrida 7, 11-21.

Barbehenn, R.V., 2001. Roles of peritrophic membranes in protecting herbivorous insects from ingested plant allelochemicals. Arch. Insect Biochem. Physiol. 47, 86-99.

Barbehenn, R.V., Martin, M.M., 1994. Tannin sensitivity in larvae of *Malacosoma disstria* (Lepidoptera): Roles of the peritrophic envelope and midgut oxidation. J. Chem. Ecol. 20, 1985-2001.

Barbehenn, R.V., Stannard, J., 2004. Antioxidant defense of the midgut epithelium by the peritrophic envelope in caterpillars. J. Insect Physiol. 50, 783-790.

Barry, M.K., Triplett, A.A., Christensen, A.C., 1999. A peritrophin-like protein expressed in the embryonic tracheae of *Drosophila melanogaster*. Insect Biochem. Mol. Biol. 29, 319-327.

Bartnicki-Garcia, S., Persson, J., Chanzy, H., 1994. An electron microscope and electron diffraction study of the effect of calcofluor and Congo red on the biosynthesis of chitin in vitro. Arch. Biochem. Biophys. 310, 6-15.

Bayes, A., Comellas-Bigler, M., Rodriguez de la Vega, M., Maskos, K., Bode, W., Aviles, F.X., Jongsma, M.A., Beekwilder, J., Vendrell, J., 2005. Structural basis of the resistance of an insect carboxypeptidase to plant protease inhibitors. Proc. Natl. Acad. Sci. USA 102, 16602-16607.

Baum, J.A., Bogaert, T., Clinton, W., Heck, G.R., Feldmann, P., Ilagan, O., Johnson, S., Plaetinck, G., Munyikwa, T., Pleau, M., Vaughn, T., Roberts, J., 2007. Control of coleopteran insect pests through RNA interference. Nat. Biotechnol. 25, 1322-1326.

Behr, M., Hoch, M., 2005. Identification of the novel evolutionary conserved obstructor multigene family in invertebrates. FEBS Lett. 579, 6827-6833.

Bell, S.L., Xu, G., Khatri, I.A., Wang, R., Rahman, S., Forstner, J.F., 2003. *N*-linked oligosaccharides play a role in disulphide-dependent dimerization of intestinal mucin Muc2. Biochem. J. 373, 893-900.

Bernays, E.A., Chamberlain, D.J., 1980. A study of tolerance of ingested tannin in *Schistocerca gregaria.* J. Insect Physiol. 26, 415-420.

Berner, R., Rudin, W., Hecker, H., 1983. Peritrophic membranes and protease activity in the midgut of the malaria mosquito, *Anopheles stephensi* (Liston) (Insecta: Diptera) under normal and experimental conditions. J. Ultrastruct. Res. 83, 195-204.

Bettencourt, R., Terenius, O., Faye, I., 2002. *Hemolin* gene silencing by ds-RNA injected into *Cecropia* pupae is lethal to next generation embryos. Insect Mol. Biol. 11, 267-271.

Beye, M., Haertel, S., Hagen, A., Hasselmann, M., Omholt, S.W., 2002. Specific developmental gene silencing in the honey bee using a homeobox motif. Insect Mol. Biol. 11, 527-532.

Billingsley, P.F., 1990. Blood digestion in the mosquito, *Anopheles stephensi* Liston (Diptera: Culicidae): Partial characterization and post feeding activity of midgut aminopeptidases. Arch. Insect Biochem. Physiol. 15, 149-163.

Bischoff, D.S., Slavicek, J.M., 1997. Molecular analysis of an *enhancin* gene in the *Lymantria dispar* nuclear polyhedrosis virus. J. Virol. 71, 8133-8140.

Blair, D.E., Hekmat, O., Schuttelkopf, A.W., Shrestha, B., Tokuyasu, K., Withers, S.G., van Aalten, D.M., 2006. Structure and mechanism of chitin deacetylase from the fungal pathogen *Colletotrichum lindemuthianum*. Biochemistry 45, 9416-9426.

Blair, D.E., Schuttelkopf, A.W., MacRae, J.I., van Aalten, D.M., 2005. Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. Proc. Natl. Acad. Sci. USA 102, 15429-15434.

Blom, N., Gammeltoft, S., Brunak, S., 1999. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol. 294, 1351-1362.

Bode, W., Gomis-Rüth, F.-X., Huber, R., Zwilling, R., Stöcker, W., 1992. Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases. Nature 358, 164-167.

Bolognesi, R., Arakane, Y., Muthukrishnan, S., Kramer, K.J., Terra, W.R., Ferreira, C., 2005. Sequences of cDNAs and expression of genes encoding chitin synthase and chitinase in the midgut of *Spodoptera frugiperda*. Insect Biochem. Mol. Biol. 35, 1249-1259.

Bolognesi, R, Ribeiro, A.F., Terra, W.R., Ferreira, C., 2001. The peritrophic membrane of *Spodoptera frugiperda*: Secretion of peritrophins and role in immobilization and recycling digestive enzymes. Arch. Insect Biochem. Physiol. 47, 62-75.

Bond, J.S., Beynon, R.J., 1995. The astacin family of metallo-endopeptidases. Protein Sci. 4, 1247-1261.

Boraston, A.B., Bolam, D.N., Gilbert, H.J., Davies, G.J., 2004. Carbohydrate-binding modules: Fine tuning polysaccharide recognition. Biochem. J. 382, 769-781.

Bown, D.P., Wilkinson, H.S., Gatehouse, J.A., 1998. Midgut carboxypeptidase from *Helicoverpa armigera* (Lepidoptera: Noctuidae) larvae: Enzyme characterization, cDNA cloning and expression. Insect Biochem. Mol. Biol. 28, 739-749.

Brady, L., Brzozowski, A.M., Derewenda, Z.S., Dodson, E., Dodson, G., Tolley, S., Turkenburg, J.P., Christiansen, L., Huge-Jensen, B., Norskov, L., Thim, L., Menge, U., 1990. A serine protease triad forms the catalytic centre of a triacylglycerol lipase. Nature 343, 767-770.

Bragatto, I., Genta, F.A., Ribeiro, A.F., Terra, W.R., Ferreira, C., 2010. Characterization of a β-1, 3-glucanase active in the alkaline midgut of *Spodoptera frugiperda* larvae and its relation to β-glucan-binding proteins. Insect Biochem. Mol. Biol. 40, 861-872.

Brandt, C., Adang, M., Spence, K., 1978. The peritrophic membrane: Ultrastructural analysis and function as a mechanical barrier to microbial infection in *Orgyia pseudotsugata*. J. Invertebr. Pathol. 32, 12-24.

Broehan, G., Zimoch, L., Wessels, A., Ertas, B., Merzendorfer, H., 2007. A chymotrypsin-like serine protease interacts with the chitin synthase from the midgut of the tobacco hornworm. J. Exp. Biol. 210, 3636-3643.

Brown, S.J., Mahaffey, J.P., Lorenzen, M.D., Denell, R.E., Mahaffey, J.W., 1999. Using RNAi to investigate orthologous homeotic gene function during development of distantly related insects. Evol. Dev. 1, 11-15.

Bucher, G.E., Bracken, G.K., 1976. The bertha armyworm, *Mamestra configurata* (Lepidoptera: Noctuidae). Artificial diet and rearing technique. Can. Entomol. 108, 1327-1338.

Bucher, G., Scholten, J., Klingler, M., 2002. Parental RNAi in *Tribolium* (Coleoptera). Curr. Biol. 5, R85-R86.

Caldeira, W., Dias, A.B., Terra, W.R., Ribeiro, A.F., 2007. Digestive enzyme compartmentalization and recycling and sites of absorption and secretion along the midgut of *Dermestes maculatus* (Coleoptera) larvae. Arch. Insect Biochem. Physiol. 64, 1-18.

Campbell, P.M., Cao, A.T., Hines, E.R., East, P.D., Gordon, K.H.J., 2008. Proteomic analysis of the peritrophic matrix from the gut of the caterpillar, *Helicoverpa armigera*. Insect Biochem. Mol. Biol. 38, 950-958.

Carrell, R.W., Boswell, D.R., 1986. Serpins: The superfamily of plasma serine proteinase inhibitors. In: Proteinase Inhibitors (ed. by Barrett, A.J., Salvesen, G.), Elsevier, New York, 403-420.

Casu, R., Eisemann, C., Pearson, R., Riding, G., East, I., Donaldson, A., Cadogan, L., Tellam, R.L., 1997. Antibody-mediated inhibition of the growth of larvae from an insect causing cutaneous myiasis in a mammalian host. Proc. Natl. Acad. Sci. USA 94, 8939-8944.

Chamankhah, M., Braun, L., Visal-Shah, S., O'Grady, M., Baldwin, D., Shi, X., Hemmingsen S.M., Alting-Mees, M., Hegedus, D.D., 2003. *Mamestra configurata* serpin-1 homologues: Cloning, localization and developmental regulation. Insect Biochem. Mol. Biol. 33, 355-369.

Chan, C.Y., Carmack, C.S., Long, D.D., Maliyekkel, A., Shao, Y., Roninson, I.B., Ding, Y., 2009. Structural interpretation of the effect of GC-content on efficiency of RNA interference BMC Bioinformatics 10, S33.

Chapman, R.F., 1985. Structure of the digestive system. In: Comprehensive Insect Physiology, Biochemistry and Pharmacology (ed. by Kerkut, G.A., Gilbert, L.I.), Vol: 4, Pergamon, Oxford, 121-165.

Chapman, G.W., 1987. A proteinaceous competitive inhibitor of lipase isolated from *Helianthus annuus* seeds. Phytochemistry 26, 3127-3131.

Chen, J., Brown, M.R., Hua, G., Adang, M.J., 2005. Comparison of the localization of *Bacillus thuringiensis* Cry1A delta-endotoxins and their binding proteins in larval midgut of tobacco hornworm, *Manduca sexta*. Cell Tissue Res. 321, 123-129.

Cho, Y.W., Jang, J., Park, C.R., Ko, S.W., 2000. Preparation and solubility in acid and water of partially deacetylated chitins. Biomacromolecules 1, 609-614.

Chou, H.H., Holmes, M.H., 2001. DNA sequence quality trimming and vector removal. Bioinformatics 17, 1093-1104.

Christeller, J.T., Laing, W.A., Markwick, N.P., Burgess, E.P.J., 1992. Midgut protease activities in 12 phytophagous lepidopteran larvae: Dietary and protease inhibitor interactions. Insect Biochem. Mol. Biol. 22, 735-746.

Christeller, J.T., Poulton, J., Markwick, N.M., Simpson, R.M., 2010. The effect of diet on the expression of lipase genes in the midgut of the light brown apple moth (*Epiphyas postvittana* Walker; Tortricidae). Insect Mol. Biol. 19, 9-25.

Church, G.M., Gilbert, W., 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA 81, 1991-1995.

Cohen, E., Casida, J.E., 1980. Properties of *Tribolium* gut chitin synthetase. Pestic. Biochem. Physiol. 13, 121-128.

Cohen, A.L., Marlow, D.P., Garner, G.E., 1968. A rapid critical point method using fluorocarbons (“Freons”) as intermediate and transitional fluids. J. Microsc. 7, 331-342.

Coutinho, P.M., Henrissat, B., 1999. Carbohydrate-active enzymes: An integrated database approach. In: Recent Advances in Carbohydrate Bioengineering (ed. by Gilbert, H.J., Davies, G., Henrissat, H., Svensson, B.), The Royal Society of Chemistry, Cambridge, 3-12.

Czapla, T.H., Lang, B.A., 1990. Effect of plant lectins on the larval development of European corn borer (Lepidoptera: Pyralidae) and southern corn rootworm (Coleoptera: Chrysomelidae). J. Econ. Entomol. 83, 2480-2485.

Daves, C.A., Williams, W.P., Davis, F.M., Baker, G.T., Ma, P.W.K., Monroe, W.A., Mohan, S., 2007. Plant resistance and its effect on the peritrophic membrane of southwestern corn borer (Lepidoptera: Crambidae) larvae. J. Econ. Entomol. 100, 976-983.

Del Rincon-Castro, M.C., Ibarra, J.E., 2005. Effect of a nucleopoyhedrovirus of *Autographa californica* expressing the enhancin gene of *Trichoplusia ni* granulovirus on *T. ni* larvae. Biocontrol Sci. Techn. 15, 701-710.

Derksen, A.C.G., Granados, R.R., 1988. Alteration of a lepidopteran peritrophic membrane by baculoviruses and enhancement of viral infectivity. Virology 167, 242-250.

Devenport, M., Alvarenga, P.H., Shao, L., Fujioka, H., Bianconi, M.L., Oliveira, P.L., Jacobs-Lorena, M., 2006. Identification of the *Aedes aegypti* peritrophic matrix protein AeIMUCI as a heme-binding protein. Biochemistry 45, 9540-9549.

Devenport, M., Fujioka, H., Donnelly-Doman, M., Shen, Z., Jacobs-Lorena, M., 2005. Storage and secretion of Ag-Aper14, a novel peritrophic matrix protein, and Ag-Muc1 from the mosquito *Anopheles gambiae*. Cell Tissue Res. 320, 175-185.

Devenport, M., Fujioka, H., Jacobs-Lorena, M., 2004. Storage and secretion of the peritrophic matrix protein Ag-Aper1 and trypsin in the midgut of *Anopheles gambiae*. Insect Mol. Biol. 13, 349-358.

Devine, P.L., McKenzie, F.C., 1992. Mucins: Structure, function and association with malignancy. BioEssays 14, 619-625.

Dimopoulos, G., Seeley, D., Wolf, A., Kafatos, F.C., 1998. Malaria infection of the mosquito *Anopheles gambiae* activates immune-responsive genes during critical transition stages of the parasite life cycle. EMBO J. 21, 6115-6123.

Dinglasan, R.R., Devenport, M., Florens, L., Johnson, J.R., McHugh, C.A., Donnelly-Doman, M., Carucci, D.J., Yates, J.R., Jacobs-Lorena, M., 2009. The *Anopheles gambiae* adult midgut peritrophic matrix proteome. Insect Biochem. Mol. Biol. 39, 125-134.

Dixit, R., Arakane, Y., Specht, C.A., Richard, C., Kramer, K.J., Beeman, R.W., Muthukrishnan, S., 2008. Domain organization and phylogenetic analysis of proteins from the chitin deacetylase gene family of *Tribolium castaneum* and three other species of insects. Insect Biochem. Mol. Biol. 38, 440-451.

Dong, D-J., He, H-J., Chai, L-Q., Wang, J-X., Zhao, X-F., 2007. Identification of differentially expressed genes during larval molting and metamorphosis of *Helicoverpa armigera*. BMC Dev. Biol. 7, 73.

Dorner, R., Peters, W., 1988. Localization of sugar components of glycoproteins in peritrophic membranes of larvae of Diptera (Culicidae, Simuliidae). Entomol. Gen. 14, 11-24.

Dow, J.A.T., O'Donnell, M.J., 1990. Reversible alkalinization by *Manduca sexta* midgut. J. Exp. Biol. 150, 247-256.

Dunaevsky, Y.E., Elpidina, E.N., Vinokurov, K.S., Belozersky, M.A., 2005. Protease inhibitors in improvement of plant resistance to pathogens and insects. Mol. Biol. 39, 608-613.

Eaton, B.A., Fetter, R.D., Davis, G.W., 2002. Dynactin is necessary for synapse stabilization. Neuron 34, 729-741.

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792-1797.

Edwards, M.J., Moskalyk, L.A., Donelly-Doman, M., Vlaskova, M., Noriega, F.G., Walker, V.K., Jacobs-Lorena, M., 2000. Characterization of a carboxypeptidase A gene from the mosquito *Aedes aegypti*. Insect Mol. Biol. 9, 33-38.

Eguchi, M., Iwamoto, A., Yamaguhi, K., 1982. Interaction of proteases from the midgut lumen, epithelial and peritrophic membrane of the silkworm *Bombyx mori* l. Comp. Biochem. Physiol. 72A, 359-363.

Eisemann, C.H., Donaldson, R.A., Pearson, R.D., Cadogan, L.C., Vuocolo, T., Tellam, R.L., 1994. Larvicidal activity of lectins on *Lucilia cuprina*: Mechanism of action. Entomol. Exp. Appl. 72, 1-10.

Eisemann, C.H., Wijffels, G., Tellam, R.L., 2001. Secretion of the type 2 peritrophic matrix protein, peritrophin-15, from the cardia. Arch. Insect Biochem. Physiol. 47, 76-85.

Eisenhaber, B., Bork, P., Yuan, Y., Loeffler, G., Eisenhaber, F., 2000. Automated annotation of GPI anchor sites: Case study *C. elegans*. Trends Biochem. Sci. 25, 340-341.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., Tuschl, T., 2001. Duplexes of 21 nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411, 494-498.

Elvin, C., Vuocolo, T., Pearson, R., East, I.J., Riding, G., Eisemann, C., Tellam, R.L., 1996. Characterization of a major peritrophic membrane protein, peritrophin-44, from the larvae of *Lucilia cuprina*: cDNA and deduced amino acid sequences. J. Biol. Chem. 271, 8925-8935.

Emmenlauer, M., Ronneberger, O., Ponti, A., Schwarb, P., Griffa, A., Filippi, A., Nitschke, R., Driever, W., Burkhardt, H., 2009. XuvTools: Free, fast and reliable stitching of large 3D datasets. J. Microsc. 233, 42-60.

Erlandson, M., 1990. Biological and biochemical comparison of *Mamestra configurata* and *Mamestra brassica* nuclear polyhedrosis virus isolates pathogenic for the bertha armyworm, *Mamestra configurata* (Lepidoptera: Noctuidae). J. Invertebr. Pathol. 56, 47-56.

Erlandson, M., Hegedus, D., Baldwin, D., Noakes, A., Toprak, U., 2010. Characterization of the *Mamestra configurata* (Lepidoptera: Noctuidae) larval midgut protease complement and adaptation to feeding on artificial diet, *Brassica* species and protease inhibitor. Arch. Insect Biochem. Physiol. 75, 70-91.

Erlandson, M., Newhouse, S., Moore, K., Janmaat, A., Myers, J., Theilmann, D., 2007. Characterization of baculovirus isolates from *Trichoplusia ni* populations from vegetable greenhouses. Biol. Control 41, 256-263.

Fang, S., Wang, L., Guo, W., Zhang, X., Peng, D., Luo, C., Yu, Z., Sun, M., 2009. *Bacillus thuringiensis* bel protein enhances the toxicity of Cry1Ac protein to *Helicoverpa armigera* larvae by degrading insect intestinal mucin. Appl. Environ. Microbiol. 75, 5237-5243.

Fernandez, L.E., Aimanova, K.G., Gill, S.S., Bravo, A., Soberon, M., 2006. A GPI-anchored alkaline phosphatase is a functional midgut receptor of Cry11Aa toxin in *Aedes aegypti* larvae. Biochem. J. 394, 77-84.

Ferreira, A.H.P., Cristofoletti, P.T., Lorenzinic, D.M., Guerra, L.O., Paiva, P.B., Briones, M.R.S., Terra, W.R., Ferreira, C., 2007. Identification of midgut microvillar proteins from *Tenebrio molitor* and *Spodoptera frugiperda* by cDNA library screenings with antibodies. J. Insect Physiol. 53, 1112-1124.

Ferreira, C., Capella, A.N., Sitnik, R., Terra, W.R., 1994. Digestive enzymes in midgut cells, endo- and ectoperitrophic contents, and peritrophic membranes of *Spodoptera frugiperda* (Lepidoptera) larvae. Arch. Insect Biochem. Physiol. 26, 299-313.

Ferreira, C., Ribeiro, A.F., Terra, W.R., 1981. Fine structure of the larval midgut of the fly *Rhynchosciara* and its physiological implications. J. Insect Physiol. 27, 559-570.

Filho, B.P.D., Lemos, F.J.A., Secundino, N.F.C., Pascoa, V., Pereira, S.T., Pimenta, P.F.P., 2002. Presence of chitinase and beta-N-acetylglucosaminidase in the *Aedes aegypti*: A chitinolytic system involving peritrophic matrix formation and degradation. Insect Biochem. Mol. Biol. 32, 1723-1729.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C., 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391, 806-811.

Fitches, E., Wilkerson, H., Bell, H., Bown, D.P., Gatehouse, J.A., Edwards, J.P., 2004. Cloning, expression and functional characterization of chitinase from larvae of tomato moth (*Lacanobia oleracea*): A demonstration of the insecticidal activity of insect chitinase. Insect Biochem. Mol. Biol. 34, 1037-1050.

Forstner, J.F., Forstner, G.G., 1994. Gastrointestinal mucus. In: Physiology of the Gastrointestinal Tract (ed. by Johnson, L.R.), Raven Press, New York, 1255-1283.

Freyvogel, T.A., Jaquet, C., 1965. The prerequisites for the formation of the peritrophic membrane in Culicidae females. Acta Trop. 22, 148-154.

Fukamizo, T., Kramer, K.J., 1985. Mechanism of chitin hydrolysis by the binary chitinase system in insect moulting fluid. Insect Biochem. 15, 141-145.

Furusawa, T., Takayama, E., Ishihara, R., Hayashi, Y., 1993. Double stranded ribonuclease activity in the digestive juice and midgut of the silkworm, *Bombyx mori*. Comp. Biochem. Physiol. 104B, 795-801.

Gaines, P.J., Walmsley, S.J., Wisnewski, N., 2003. Cloning and characterization of five cDNAs encoding peritrophin-A domains from the cat flea, *Ctenocephalides felis*. Insect Biochem. Mol. Biol. 33, 1061-1073.

Gagou, M.E., Kapsetaki, M., Turberg, A., Kafetzopoulos, D., 2002. Stage specific expression of the chitin synthase *DmeChSA* and *DmeChSB* genes during the onset of *Drosophila* metamorphosis. Insect Biochem. Mol. Biol. 32, 141-146.

Gallo, L.G., Corsaro, B.G., Hughes, P.R., Granados, R.R., 1991. *In vivo* enhancement of baculovirus infection by the viral enhancing factor of a granulosis virus of the cabbage looper, *Trichoplusia ni* (Lepidoptera: Noctuidae). J. Invertebr. Pathol. 58, 203-210.

Galloway, C.S., Wang, P., Winstanley, D., Jones, I.M., 2005. Comparison of the bacterial enhancin-like proteins from *Yersinia* and *Bacillus* spp. with a baculovirus enhancin. J. Invertebr. Pathol. 90, 134-137.

Garcia-Zapien, A.G., Gonzalez-Robles, A., Mora-Galindo, J., 1999. Congo red effect of cyst viability and cell wall structure of encysting *Entamoeba invadens*. Arch. Med. Res. 30, 106-115.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. In: The Proteomics Protocols Handbook (ed. by Walker, J.M.), Humana Press, Totowa, USA, 571-607.

Geldhof, P., Murray, L., Couthier, A., Gilleard, J.S., McLauchlan, G., Knox, D.P., Britton, C., 2006. Testing the efficacy of RNA interference in *Haemonchus contortus*. Int. J. Parasitol. 36, 801-810.

Genta, F.A., Blanes, L., Cristofoletti, P.T., do Lago, C.L., Terra, W.R., Ferreira, C., 2006. Purification, characterization and molecular cloning of the major chitinase from *Tenebrio molitor* larval midgut. Insect Biochem. Mol. Biol. 36, 789-800.

Genta, F.A., Bragatto, I., Terra, W.R., Ferreira, C., 2009. Purification, characterization and sequencing of the major β-1, 3-glucanase from the midgut of *Tenebrio molitor* larvae. Insect Biochem. Mol. Biol. 39, 861-874.

Genta, F.A., Terra, W.R., Ferreira, C., 2003. Action pattern, specificity, lytic activities, and physiological role of five digestive β-glucanases isolated from *Periplaneta americana*. Insect Biochem. Mol. Biol. 33, 1085-1097.

Gillott, C., 2005. Entomology. Third Edition. Springer, Dordrecht, The Netherlands.

Girard, C., Jouanin, L., 1999. Molecular cloning of a gut-specific chitinase cDNA from the beetle *Phaedon cochleariae*. Insect Biochem. Mol. Biol. 29, 549-556.

Gomis-Rüth, F.-X., Stöcker, W., Huber, R., Zwilling, R., Bode, W., 1993. Refined 1.8 A° X-ray crystal structure of astacin, a zinc-endopeptidase from the crayfish *Astacus astacus* L. Structure determination, refinement, molecular structure and comparison with thermolysin. J. Mol. Biol. 229, 945-968.

Granados, R.R., Fu, Y., Corsaro, B., Hughes, P.R., 2001. Enhancement of *Bacillus thuringiensis* toxicity to lepidopterous species with the enhancin from *Trichoplusia ni* granulovirus. Biol. Control 20, 153-159.

Grossi de Sa, M.F., Chrispeels, M.J., 1997. Molecular cloning of bruchid (*Zabrotes subfasciatus*) alpha-amylase cDNA and interactions of the expressed enzyme with bean amylase inhibitors. Insect Biochem. Mol. Biol. 27, 271-281.

Guo, H-F., Fang, J-C., Liu, B-S., Wang, J-P., Zhong, W-F., Wan, F-H., 2007a. Enhancement of the biological activity of nucleopolyhedrovirus through disruption of the peritrophic matrix of insect larvae by chlorfluazuron. Pest Manag. Sci. 63, 68-74.

Guo, H-F., Fang, J-C., Wang, J-P., Zhong, W-F., Liu, B-S., 2007b. Interaction of *Xestia c-nigrum* granulovirus with peritrophic matrix and *Spodoptera litura* nucleopolyhedrovirus in *Spodoptera litura*. J. Econ. Entomol. 100, 20-25.

Guo, W., Li, G,. Pang, Y., Wang, P., 2005. A novel chitin-binding protein identified from the peritrophic membrane of the cabbage looper, *Trichoplusia ni*. Insect Biochem. Mol. Biol. 35, 1224-1234.

Hammond, S.M., Bernstein, E., Beach, D., Hannon, G.J., 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. Nature 404, 293-296.

Harper, M.S., Granados, R.R., 1999. Peritrophic membrane structure and formation of larval *Trichoplusia ni* with an investigation on the secretion patterns of a PM mucin. Tissue&Cell 31, 202-211.

Harper, M.S., Hopkins, T.L., 1997. Peritrophic membrane structure and secretion in European corn borer larvae (*Ostrinia nubilalis*). Tissue and Cell 29, 463-475.

Harper, M.S., Hopkins, T.L., Czapla, T.H., 1998. Effect of wheat germ agglutinin on formation and structure of the peritrophic membrane in European corn borer (*Ostrinia nubilalis*) larvae. Tissue and Cell 30, 166-176.

Hashimoto, Y., Corsaro, B.G., Granados, R.R., 1991. Location and nucleotide sequence of the gene encoding the viral enhancing factor of the *Trichoplusia ni* granulosis virus. J. Gen. Virol. 72, 2645-2651.

Hayakawa, T., Shitomi, Y., Miyamoto, K., Hori, H., 2004. GalNAc pretreatment inhibits trapping of *Bacillus thuringiensis* Cry1Ac on the peritrophic membrane of *Bombyx mori* FEBS Lett. 576, 331-335.

He, N., Botelho, J.M., McNall, R.J., Belozerov, V., Dunn, W.A., Mize, T., Orlando, R., Willis, J.H., 2007. Proteomic analysis of cast cuticles from *Anopheles gambiae* by tandem mass spectrometry. Insect Biochem. Mol. Biol. 37, 135-146.

Hegedus, D., Baldwin, D., O'Grady, M., Braun, L., Gleddie, S., Sharpe, A., Lydiate, D., Erlandson, M., 2003. Midgut proteases from *Mamestra configurata* (Lepidoptera: Noctuidae) larvae: Characterization, cDNA cloning and expressed sequence tag analysis. Arch. Insect Biochem. Physiol. 53, 30-47.

Hegedus, D., Erlandson, M., Baldwin, D., Hou, X., Chamankhah, M., 2008. Differential expansion and evolution of the exon family encoding the serpin-1 reactive centre loop has resulted in divergent serpin repertoires among the Lepidoptera. Gene 418, 15-21.

Herrero, S., Ansems, M., Van Oers, M.M., Vlak, J.V., Bakker, P.L., de Maagd, R.A., 2007. REPAT, a new family of proteins induced by bacterial toxins and baculovirus infection in *Spodoptera exigua*. Insect Biochem. Mol. Biol. 37, 1109-1118.

Herth, W., 1980. Calcofluor white and Congo red inhibit chitin microfibril assembly in poteriochromonas: Evidence for a gap between polymerization and microfibril formation. J. Cell. Biol. 3, 567-569.

Hide, W.A., Chan, L., Li, W.H., 1992. Structure and evolution of the lipase superfamily. J. Lipid Res. 33, 167-178.

Hogenkamp, D.G., Arakane, Y., Zimoch, L., Merzendorfer, H., Kramer, K.J., Beeman, R.W., Kanost, M.R., Specht, C.A., Muthukrishnan, S., 2005. Chitin synthase genes in *Manduca sexta*: Characterization of a gut-specific transcript and differential tissue expression of alternately spliced mRNAs during development. Insect Biochem. Mol. Biol. 35, 529-540.

Holt, D.C., Fischer, K., Allen, G.E., Wilson, D., Wilson, P., Slade, R., Currie, B.J., Walton, S.F., Kemp, D.J., 2003. Mechanisms for a novel immune evasion strategy in the scabies mite *Sarcoptes scabiei*: A multigene family of inactivated serine proteases. J. Investig. Dermatol. 121, 1419-1424.

Hooper, N.M., 1994. Families of zinc metalloproteases. FEBS Lett. 354, 1-6.

Hoover, K., Humphries, M.A., Gendron, A.R., Slavicek, J.M., 2010. Impact of viral enhancin genes on potency of Lymantria dispar multiple nucleopolyhedrovirus in *L. dispar* following disruption of the peritrophic matrix. J. Invertebr. Pathol. 104, 150-152.

Hopkins, T.L., Harper, M.S., 2001. Lepidopteran peritrophic membranes and effects of dietary wheat germ agglutinin on their formation and structure Arch. Insect Biochem. Physiol. 47, 100-109.

Huang, X., Madan, A., 1999. CAP3: A DNA sequence assembly program. Genome Res. 9, 868-877.

Huang, X., Zhang, H., Zen, K.C., Muthukrishnan, S., Kramer, K.J., 2000. Homology modeling of the insect chitinase catalytic domain-oligosaccharide complex and the role of a putative active site tryptophan in catalysis. Insect Biochem. Mol. Biol. 30, 107-117.

Huber, M., Cabib, E., Miller, L.H., 1991. Malaria parasite chitinase and penetration of the mosquito peritrophic membrane. Proc. Natl. Acad. Sci. USA 88, 2807-2810.

Ibrahim, G.H., Smartt, C.T., Kiley, L.M., Christensen, B.M., 2000. Cloning and characterization of a chitin synthase cDNA from the mosquito *Aedes aegypti*. Insect Biochem. Mol. Biol. 30, 1213-1222.

Jacobs-Lorena, M., Oo, M.M., 1996. The peritrophic membrane of insects. In: Biology of Disease Vectors: A Molecular, Physiological, and Populational Approach (ed. by Beaty, B., Marqquardt, W.), University Press of Colorado, Denver, 318-332.

Jakubowska, A.K., Caccia, S., Gordon, K.H., Ferre, J., Herrero, S., 2010. Down-regulation of a chitin deacetylase-like protein in response to baculovirus infection and its application for improving baculovirus infectivity. J. Virol. 84, 2547-2555.

Janecek, S., 1997. Alpha-amylase family: Molecular biology and evolution. Prog. Biophys. Mol. Biol. 67, 67-97.

Jang, M-K., Kong, B-G., Jeong, Y-I., Lee, C.H., Nah, J-W., 2004. Physicochemical characterization of α-chitin, β-chitin, and γ-chitin separated from natural resources. J. Polymer Sci. 42A, 3423-3432.

Jasrapuria, S., Arakane, Y., Osman, G., Kramer, K.J., Beeman, R.W., Muthukrishnan, S., 2010. Genes encoding proteins with peritrophin A-type chitin-binding domains in *Tribolium castaneum* are grouped into three distinct families based on phylogeny, expression and function. Insect Biochem. Mol. Biol. 40, 214-227.

Jehle, J.A., Blissard, G.W., Bonning, B.C., Cory, J.S., Herniou, E.A., Rohrmann, G.F., Theilmann, D.A., Thiem, S.M., Vlak, J.M., 2006. On the classification and nomenclature of baculoviruses: A proposal for revision. Arch. Virol. 151, 1257-1266.

Jiang, Q., Hall, M., Noriega, F.G., Wells, M., 1997. cDNA cloning and pattern of expression of an adult, female-specific chymotrypsin from *Aedes aegypti* midgut. Insect Biochem. Mol. Biol. 27, 283-289.

Jochim, R.C., Teixeira, C.R., Laughinghouse, A., Mu, J., Oliveira, F., Gomes, R.B., Elnaiem, D-E., Valenzuela, J.G., 2008. The midgut transcriptome of *Lutzomyia longipalpis*: Comparative analysis of cDNA libraries from sugar-fed, blood-fed, postdigested and *Leishmania infantum chagasi* infected sand flies. BMC Genomics 9, 15.

Jongeneel, C.V., Bouvier, J., Bairoch, A., 1989. A unique signature identifies a family of zinc-dependent metallopeptidases. FEBS Lett. 242, 211-214.

Julenius, K., Mølgaard, A., Gupta, R., Brunak, S., 2005. Prediction, conservation analysis and structural characterization of mammalian mucin-type *O*-glycosylation sites. Glycobiology 15, 153-164.

Jurat-Fuentes, J.L., Adang, M.J., 2004. Characterization of a Cry1Ac receptor alkaline phosphatase in susceptible and resistant *Heliothis virescens* larvae. Eur. J. Biochem. 271, 3127-3135.

Kaiser-Alexnat, R., 2009. Protease activities in the midgut of western corn rootworm (*Diabrotica virgifera virgifera* LeConte). J. Invertebr. Pathol. 100, 169-174.

Karlsson, N.G., Olson, F.J., Joval, P.A., Andersch, Y., Enerback, L., Hansson, G.C., 2000. Identification of transient glycosylation alterations of sialylated mucin oligosaccharides during infection by the rat intestinal parasite *Nippostrongylus brasiliensis*. Biochem. J. 350, 805-814.

Kato, N., Muller, C.R., Fuchs, J.F., Wessely, V., Lan, Q., Christensen, B.M., 2006. Regulatory mechanisms of chitin biosynthesis and role of chitin in peritrophic matrix formation in the midgut of adult *Aedes aegypti*. Insect Biochem. Mol. Biol. 36, 1-9.

Kawabata, S., Nagayama, R., Hirata, M., Shigenaga, T., Agarwala, K.L., Saito, T., Cho, J., Nakajima, H., Takagi, T., Iwanaga, S., 1996. Tachycitin, a small granular component in horseshoe crab hemocytes, is an antimicrobial protein with chitin-binding activity. J. Biochem. 120, 1253-1260.

Kennerdell, J.R., Carthew, R.W., 1998. Use of dsRNA mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. Cell 95, 1017-1026.

Khajuria, C., Buschman, L.L., Chen, M-S., Muthukrishnan, S., Zhu, K.Y., 2010. A gut-specific chitinase gene essential for regulation of chitin content of peritrophic matrix and growth of *Ostrinia nubilalis* larvae. Insect Biochem. Mol. Biol. 40, 621-629.

Kim, M.G., Shin, S.W., Bae, K.S., Kim, S.C., Park, H.Y., 1998. Molecular cloning of chitinase cDNAs from the silkworm, *Bombyx mori* and the fall webworm, *Hyphantria cunea.* Insect Biochem. Mol. Biol. 28, 163-171.

Koga, K., Hara, A., Akune, S., 1969. Correlation of nucleolytic activities with the stage of growth of the silkworm. J. Insect Physiol. 15, 799-806.

Koga, D., Fujimoto, H., Funakoshi, T., Utsumi, T., Ide, A., 1989. Appearance of chitinolytic enzymes in integument of *Bombyx mori* during the larval-pupal transformation. Evidence for zymogenic forms. Insect Biochem. 19, 123-128.

Koles, K., Lim, J-M., Aoki, K., Porterfield, M., Tiemeyer, M., Wells, L., Panin, V., 2007. Identification of N-glycosylated proteins from the central nervous system of *Drosophila melanogaster*. Glycobiology 17, 1388-1403.

Konno, K., Hirayama, C., Nakamura, M., Tateishi, K., Tamura, Y., Hattori, M., Kohno, K., 2004. Papain protects papaya trees from herbivorous insects: role of cysteine proteases in latex. Plant J. 37, 370-378.

Kramer, K.J., Koga, D., 1986. Insect chitin: Physical state, synthesis, degradation and metabolic regulation. Insect Biochem. 16, 851-877.

Kramer, K.J., Muthukrishnan, S., 1997. Insect chitinases: Molecular biology and potential use as biopesticides. Insect Biochem. Mol. Biol. 27, 887-900.

Kramer, K.J., Muthukrishnan, S., 2005. Chitin metabolism in insects. In: Comprehensive Molecular Insect Science (ed. by Gilbert, L.I., Iatrou, K., Gill, S.S.), Vol:4, Elsevier, New York, 111-144.

Kramer, K.J., Corpuz, L., Choi, H.K., Muthukrishnan, S., 1993. Sequence of a cDNA and expression of the gene encoding epidermal and gut chitinases of *Manduca sexta*. Insect Biochem. Mol. Biol. 23, 691-701.

Kraulis, J., Clore, G.M., Nilges, M., Jones, T.A., Pettersson, G., Knowles, J., Gronenborn, A.M., 1989. Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei.* A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. Biochemistry 28, 7241-7257.

Krem, M.M., Rose, T., Di Cera, E., 2000. Sequence determinants of function and evolution in serine proteases. Trends Cardiovasc. Med. 10, 171-176.

Kumar, N.S., Tang, B., Chen, X., Tian, H., Zhang, W., 2008. Molecular cloning, expression pattern and comparative analysis of chitin synthase gene B in *Spodoptera exigua*. Comp. Biochem. Phys. 149B, 447-453.

Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105-132.

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680-685.

Lagermalm, G., Philip B., Gralenn, N., 1950. Occurrence of a network in the excrement from the larva of the clothes moth. Nature 166, 484-485.

Lapointe, R., Popham, H.J.R., Straschil, U., Goulding, D., O'Reilly, D.R., Olszewski, J.A., 2004. Characterization of two *Autographa californica* nucleopolyhedrovirus proteins, Ac145 and Ac150, which affect oral infectivity in a host-dependent manner. J. Virol. 78, 6439-6448.

Lee, R.F., 1968. The histology and histochemistry of the anterior midgut of *Periplaneta americana* (Dictyoptera: Blattidae) with reference to the formation of the peritrophic membrane. Proc. R. Entomol. Soc. Lond. Ser. 43A, 122-134.

Lee, M.J., Anstee, J.H., 1995. Endoproteases from the midgut of larval *Spodoptera litteralis* include a chymotrypsin-like enzyme with an extended binding site. Insect Biochem. Mol. Biol. 25, 49-61.

Lehane, M.J., 1997. Peritrophic matrix structure and function. Annu. Rev. Entomol. 42, 525-550.

Lehane, M.J., Msangi, A.R., 1991. Lectin and peritrophic membrane development in the gut of *Glossina m.morsitans* and a discussion of their role in protecting the fly against trypanosome infection. Med. Vet. Entomol. 5, 495-501.

Lehane, M.J., Müller, H.M., Crisanti, A., 1996. Mechanisms controlling the synthesis and secretion of digestive enzymes in insects. In: Biology of the Insect Midgut (ed. by Lehane M.J., Billingsley, P.F.), Chapman&Hall, Cambridge, 195-205.

Lehane, M.J., Aksoy, S., Gibson, W., Kerhornou, A., Berriman, M., Hamilton, J., Soares, M.B., Bonaldo, M.F., Lehane, S., Hall, N., 2003. Adult midgut expressed sequence tags from the tsetse fly *Glossina morsitans morsitans* and expression analysis of putative immune response genes. Genome Biol. 4, R63.

Lepore, L.S., Roelvink, P.R., Granados, R.R., 1996. Enhancin, granulosis virus protein that facilitates nucleopolyhedrovirus (NPV) infections, is a metalloprotease. J. Invertebr. Pathol. 68, 131-140.

Lertcanawanichakul, M., Wiwat, C., Bhumiratana, A., Dean, D.H., 2004. Expression of chitinase-encoding genes in *Bacillus thuringiensis* and toxicity of engineered *B. thuringiensis* subsp. *aizawa* toward *Lymantria dispar* larvae. Curr. Microbiol. 48, 175-181.

Levy, S.M., Falleiros, A.M.F., Moscardi, F., Gregorio, E.A., 2007. Susceptibility/resistance of *Anticarsia gemmatalis* larvae to its nucleopolyhedrovirus (AgMNPV): Structural study of the peritrophic membrane. J. Invertebr. Pathol. 96, 183-186.

Li, Q., Donly, C., Li, L., Willis, L.G., Theilmann, D.A., Erlandson, M., 2002a. Sequence and organization of the *Mamestra configurata* nucleopolyhedrovirus genome. Virology 294, 106-121.

Li, W., Petersen, R.A., Schuler, M.A., Berenbaum, M.R., 2002b. CYP6B cytochrome p450 monooxygenases from *Papilio canadensis* and *Papilio glaucus*: Potential contributions of sequence divergence to host plant associations. Insect Mol. Biol. 11, 543-551.

Li, Q., Li, L., Moore, K., Donly, C., Theilmann, D.A., Erlandson, M., 2003. Characterization of *Mamestra configurata* nucleopolyhedrovirus enhancin and its functional analysis via expression in an *Autographa californica* M nucleopolyhedrovirus recombinant. J. Gen. Virol. 84, 123-132.

Liao, C., Trowell, S.C., Akhurst, R.J., 1996. Binding of *Bacillus thuringiensis* insecticidal crystal proteins to the cotton bollworm, *Helicoverpa armigera* and its impact to the environment. In: Proceedings of the Second Pacific Rim Conference on Biotechnology of *Bacillus thuringiensis*, Chiang Mai, Thailand, 180-192.

Links, M.G., Crosby, W.L., 2007. APED: Another portal for examining DNA. http://aped.sourceforge.net/.

Loris, R., Hamelryck, T., Bouckaert, J., Wyns, L., 1998. Legume lectin structure. Biochim. Biophys. Acta 1383, 9-36.

Lu, Y., Zen, K.C., Muthukrishnan, S., Kramer, K.J., 2002. Site-directed mutagenesis and functional analysis of active site acidic amino acid residues D142, D144 and E146 in *Manduca sexta* (tobacco hornworm) chitinase. Insect Biochem. Mol. Biol. 32, 1369-1382.

Luschnig, S., Batz, T., Armbruster, K., Krasnow, M.A., 2006. Serpentine and vermiform encode matrix proteins with chitin-binding and deacetylation domains that limit tracheal tube length in *Drosophila*. Curr. Biol. 16, 186-194.

Ma, C., Kanost, M.R., 2000. A β,1-3-glucan-recognition protein from an insect, *Manduca sexta*, agglutinates microorganisms and activates the phenoloxidase cascade. J. Biol. Chem. 275, 7505–7514.

Male, K.B., Storey, K.B., 1981. Enzyme activities and isozyme composition of triglyceride, diglyceride and monoglyceride lipases in *Periplaneta americana*, *Locusta migratoria* and *Polia adjuncta*. Insect Biochem. 11, 423-427.

Mao, Y.B., Cai, W.J., Wang, J.W., Hong, G.J., Tao, X.Y., Wang, L.J., Huang, Y.P., Chen, X.Y., 2007. Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. Nat. Bioctechnol. 25, 1307-1313.

Marchler-Bauer, A., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Lu, S., Marchler, G.H., Mullokandov, M., Song, J.S., Tasneem, A., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., Bryant, S.H., 2009. CDD: Specific functional annotation with the conserved domain database. Nucleic Acids Res. 37, D205-D210.

Marques-Silva, S., Serrao, J.E., Mezencio, J.M.S., 2005. Peritrophic membrane protein in the larval stingless bee *Melipona quadrifasciata anthidioides*: Immunolocalization of secretory sites. Acta Histochem. 107, 23-30.

Martinez, J.P, Falomir, M.P., Gozalbo, D., 2009. Chitin: A structural biopolysaccharide. In: Encyclopedia of Life Sciences (ELS), John Wiley & Sons Ltd, Chichester, UK, 1-10.

Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., Tuschl, T., 2002. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell 110, 563-574.

Maskel, S.M., DiCapua, R.A., 1988. Comparison of protease activity of four nuclear polyhedrosis viruses. J. Invertebr. Pathol. 52, 441-452.

Mason, P.G., Arthur, A.P., Olfert, O., Erlandson, M.A., 1998. The bertha armyworm (*Mamestra configurata*) (Lepidoptera: Noctuidae) in Western Canada. Can. Entomol. 130, 321-336.

Mauchamp, B., Royer, C., Garel, A., Jalabert, A., Da Rocha, M., Grenier, A.M., Labas, V., Vinh, J., Mita, K., Kadono, K., Chavancy, G., 2006. Polycalin (chlorophyllid A binding protein): A

novel, very large fluorescent lipocalin from the midgut of the domestin silkworm *Bombyx mori* L. Insect Biochem. Mol. Biol. 36, 623-633.

Mayer, R.T., Chen, A.C., DeLoach, J.R., 1980. Characterization of a chitin synthase from the stable fly, *Stomoxys calcitrans* (L.). Insect Biochem. 10, 549-556.

Mazumdar-Leighton, S., Babu, C.R., Bennett, J., 2000. Identification of novel serine proteinase gene transcripts in the midguts of two tropical insect pests, *Scirpophaga incertulas* (Wk.) and *Helicoverpa armigera* (Hb.). Insect Biochem. Mol. Biol. 30, 57-68.

Mercer, E.H., Day, M.F., 1952. The fine structure of the peritrophic membrane of certain insects. Biol. Bull. 103, 384-394.

Merz, R.A., Horsch, M., Nyhlen, L.E., Rast, D.M., 1999. Biochemistry of chitin synthase. Exs 87, 9-37.

Merzendorfer, H., 2006. Insect chitin synthases: A review. J. Comp. Physiol. 176B, 1-15.

Merzendorfer, H., Zimoch, L., 2003. Chitin metabolism in insects: Structure, function and regulation of chitin synthases and chitinases. J. Exp. Biol. 206, 4393-4412.

Miller, N., Lehane, M.J., 1993. Ionic environment and the permeability properties of the peritrophic membrane of *Glossina morsitans morsitans*. J. Insect Physiol. 39, 139-144.

Misquitta, L., Paterson, B.M., 1999. Targeted disruption of gene function in *Drosophila* by RNA interference (RNA-i): A role for *nautilus* in embryonic somatic muscle formation. Proc. Natl. Acad. Sci. USA 96, 1451-1456.

Mitsuhashi, W., Miyamoto, K., 2003. Disintegration of the peritrophic membrane of silkworm larvae due to spindles of an entomopoxvirus. J. Invertebr. Pathol. 82, 34-40.

Mitsuhashi, W., Kawakita, H., Murakami, R., Takemoto, Y., Saiki, T., Miyamoto, K., Wada, S., 2007. Spindles of an entomopoxvirus facilitate its infection of the host insect by disrupting the peritrophic membrane. J. Virol. 81, 4235-4243.

Mittapalli, O., Sardesai, N., Shukle, R.H., 2007. cDNA cloning and transcriptional expression of a peritrophin-like gene in the hessian fly, *Mayetiola destructor* (Say). Arch. Insect Biochem. Physiol. 64, 19-29.

Mohan, S., Ma, P.W.K., Pechan, T., Bassford, E.R., Williams, W.P., Luthe, D.S., 2006. Degradation of the *S. frugiperda* peritrophic matrix by an inducible maize cysteine protease. J. Insect Physiol. 52, 21-28.

Molnar, K., Holderith Borhegyi, N., Csikos, G., Sass, M., 2001. Distribution of serpins in the tissues of the tobacco hornworm (*Manduca sexta*) larvae: Existence of new serpins possibly encoded by a gene distinct from the serpin-1 gene. J. Insect Physiol. 47, 675-687.

Morris, K., Lorenzen, M.D., Hiromasa, Y., Tomich, J.M., Oppert, C., Elpidina, E.N., Vinokurov, K., Jurat-Fuentes, J.L., Fabrick, J., Oppert, B., 2009. *Tribolium castaneum* larval gut transcriptome and proteome: A resource for the study of the coleopteran gut. J. Proteome Res. 8, 3889-3898.

Moskalyk, L.A., Oo, M.M., Jacobs-Lorena, M., 1996. Peritrophic matrix proteins of *Anopheles gambiae* and *Aedes aegypti*. Insect Mol. Biol. 5, 261-268.

Murdock, L.L., Huesing, J.E., Nielsen, S.S., Pratt, R.C., Shade, R.E., 1990. Biological effects of plant lectins on the cowpea weevil. Phytochemistry 29, 85-89.

Nagel, G., Peters, W., 1991. Formation, properties and degradation of the peritrophic membranes of larval and adult fleshflies, *Parasarcophaga argyrostoma* (Insecta, Diptera). Zoomorphology 111, 103-111.

Nègre, V., Hotelier, T., Volkoff, A-N., Gimenez, S., Cousserans, F., Mita, K., Sabau, X., Rocher, J., López-Ferber, M., d'Alençon, E., Audant, P., Sabourault, C., Bidegainberry, V., Hilliou, F., Fournier, P., 2006. Spodobase: An EST database for the lepidopteran crop pest *Spodoptera*. BMC Bioinformatics 7, 322.

Neville, A., 1975. Biology of the arthropod cuticle. Springer-Verlag, Berlin.

Ning, C., Wua, K., Liu, C., Gao, Y., Jurat-Fuentes, J.L., Gao, X., 2010. Characterization of a Cry1Ac toxin-binding alkaline phosphatase in the midgut from *Helicoverpa armigera* (Hubner) larvae. J. Insect Physiol. 56, 666-672.

Nisole, A., Stewart, D., Bowman, S., Zhang, D., Krell, P.J., Doucet, D., Cusson, M., 2010. Cloning and characterization of a Gasp homolog from the spruce budworm, *Choristoneura fumiferana*, and its putative role in cuticle formation. J. Insect Physiol. 56, 1427-1435.

Okuno, S., Takatsuka, J., Nakai, M., Ototake, S., Masui, A., Kunimi, Y., 2003. Viral enhancing activity of various stilbene-derived brighteners for a *Spodoptera litura* (Lepidoptera: Noctuidae) nucleopolyhedrovirus. Biol. Control 26, 146-152.

Oltean, D.I., Pullikuth, A.K., Lee, H-K., Gill, S.S., 1999. Partial purification and characterization of *Bacillus thuringiensis* Cry1A toxin receptor A from *Heliothis virescens* and cloning of the corresponding cDNA. Appl. Environ. Microbiol. 65, 4760-4766.

Palli, S.R., Locke, M., 1987. Hemolymph protein synthesis by the larval midgut of an insect *Calpodes ethlius* (Lepidoptera, Hesperiidae). Insect Biochem. 17, 561-572.

Pandian, G.N., Ishikawa, T., Togashi, M., Shitomi, Y., Haginoya, K., Yamamoto, S., Nishiumi, T., Hori, H., 2008. *Bombyx mori* midgut membrane protein P252, which binds to *Bacillus thuringiensis* Cry1A, is a chlorophyllide-binding protein, and the resulting complex has antimicrobial activity. Appl. Environ. Microbiol. 74, 1324-1331.

Pascoa, V., Oliveira, P.L., Dansa-Petretski, M., Silva, J.R., Alvarenga, P.H., Jacobs-Lorena, M., Lemos, F.J.A., 2002. *Aedes aegypti* peritrophic matrix and its interaction with heme during blood digestion. Insect Biochem. Mol. Biol. 32, 517-523.

Pauchet, Y., Freitak, D., Heidel-Fischer, H.M., Heckel, D.G., Vogel, H., 2009a. Immunity or digestion. Glucanase activity in a glucan-binding protein family from Lepidoptera. J. Biol. Chem. 284, 2214-2224.

Pauchet, Y., Muck, A., Svatos, A., Heckel, D.G., Preiss, S., 2008. Mapping the larval midgut lumen proteome of *Helicoverpa armigera*, a generalist herbivorous insect. J. Proteome Res. 7, 1629-1639.

Pauchet, Y., Wilkinson, P., van Munster, M., Augustin, S., Pauron, D., ffrench-Constant, R.H., 2009b. Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela tremulae* reveals new gene families in Coleoptera. Insect Biochem. Mol. Biol. 39, 403-413.

Pauchet, Y., Wilkinson, P., Vogel, H., Nelson, D.R., Reynolds, S.E., Heckel, D.G., Ffrench-Constant, R.H., 2010. Pyrosequencing the *Manduca sexta* larval midgut transcriptome: Messages for digestion, detoxification and defence. Insect Mol. Biol. 19, 61-75.

Pechan, T., Cohen, A., Williams, W.P., Luthe, D.S., 2002. Insect feeding mobilizes a unique plant defense protease that disrupts the peritrophic matrix of caterpillars. Proc. Natl Acad. Sci. USA 99, 13319-13323.

Pelegrini, P.B., Murad, A.M., Grossi-de-Sá, M.F., Mello, L.V., Romeiro, L.A.S., Noronha, E.F., Caldas, R.A., Franco, O.L., 2006. Structure and enzyme properties of *Zabrotes subfasciatus* α-amylase. Arch. Insect Biochem. Physiol. 61, 77-86.

Pelle, R., Murphy, N.B., 1993. Northern hybridization: rapid and simple electrophoretic conditions. Nucleic Acids Res. 21, 2783-2784.

Peng, J., Zhong, J., Granados, R.R., 1999. A baculovirus enhancin alters the permeability of a mucosal midgut peritrophic matrix from lepidopteran larvae. J. Insect Physiol. 45, 159-166.

Perera, O.P., Willis, J.D., Adang, M.J., Jurat-Fuentes, J.L., 2009. Cloning and characterization of the Cry1Ac-binding alkaline phosphatases (HvALP) from *Heliothis virescens*. Insect Biochem. Mol. Biol. 39, 294-302.

Perez-Vilar, J., Hill, R.L., 1999. The structure and assembly of secreted mucins. J. Biol. Chem. 274, 31751-31754.

Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J., Quackenbush, J., 2003. TIGR Gene Indices clustering tools (TGICL): A software system for fast clustering of large EST datasets. Bioinformatics 19, 651-652.

Peters, W., 1992. Peritrophic Membranes. In: Zoophysiology (ed. by Bradshaw, D., Burggren, W., Heller, H.C., Ishii, S., Langer, H., Neuweiler, G., Randall, D.J.), Vol: 130, Springer-Verlag, Berlin, 1-238.

Peters, W., Wiese, B., 1986. Permeability of the peritrophic membranes of some Diptera to labelled dextrans. J. Insect Physiol. 32, 43-49.

Peters, W., Heitmann, S., D'Haese, J., 1979. Formation and fine structure of peritrophic membranes in the earwig, *Forficula auricularia* (Dermaptera: Forficulidae). Entomol. Gen. 5, 241-254.

Peters, W., Kolb, H., Kolb-Bachofen, V., 1983. Evidence for a sugar receptor (lectin) in the peritrophic membrane of the blowfly larva, C*alliphora erythrocephala* Mg. (Diptera). J. Insect Physiol. 29, 275-280.

Piggot, C.R., Ellar, D.J., 2007. Role of receptors in *Bacillus thuringiensis* crystal toxin activity. Microbiol. Mol. Biol. Rev. 71, 255-281.

Pistillo, D., Manzi, A., Tino, A., Pilo Boyl, P., Graziani, F., Malva, C., 1998. The *Drosophila melanogaster* lipase homologs: A gene family with tissue and developmental specific expression. J. Mol. Biol. 276, 877-885.

Pitari, G., Malergue, F., Martin, F., Philippe, J.M., Massucci, M.T., Chabret, C., Maras, B., Duprè, S., Naquet, P., Galland, F., 2000. Pantetheinase activity of membrane-bound Vanin-1: Lack of free cysteamine in tissues of Vanin-1 deficient mice. FEBS Lett. 483, 149-54.

Plymale, R., Grove, M.J., Cox-Foster, D., Ostiguy, N., Hoover, K., 2008. Plant-mediated alteration of the peritrophic matrix and baculovirus infection in lepidopteran larvae. J. Insect Physiol. 54, 737-749.

Ponnuvel, K.M., Nakazawa, H., Furukawa, S., Asaoka, A., Ishibashi, J., Tanaka, H., Yamakawa, M., 2003. A lipase isolated from the silkworm *Bombyx mori* shows antiviral activity against nucleopolyhedrovirus. J. Virol. 77, 10725-10729.

Popham, H.J., Bischoff, D.S., Slavicek, J.M., 2001. Both *Lymantria dispar* nucleopolyhedrovirus enhancin genes contribute to viral potency. J. Virol. 75, 8639-8648.

Pytelkova, J., Hubert, J., Lepsik, M., Sobotnik, J., Sindelka, R., Krizkova, I., Horn, M., Mares, M., 2009. Digestive α-amylases of the flour moth *Ephestia kuehniella* - adaptation to alkaline environment and plant inhibitors. FEBS J. 276, 3531-3546.

Rajagopal, R., Sivakumar, S., Agrawal, N., Malhotra, P., Bhatnagar, R.K., 2002. Silencing of midgut aminopeptidase N of *Spodoptera litura* by double-stranded RNA establishes its role as *Bacillus thuringiensis* toxin receptor. J. Biol. Chem. 277, 46849-46851.

Ramalho-Ortigao, J.M., Jochim, R.C., Anderson, J.M., Lawyer, P.G., Pham, V-M., Kamhawi, S., Valenzuela, J.G., 2007. Exploring the midgut transcriptome of *Phlebotomus papatasi:* Comparative analysis of expression profiles of sugar-fed, blood-fed and *Leishmania major*-infected sandflies. BMC Genomics 8, 300.

Ramalho-Ortigao, J.M., Traub-Cseko, Y.M., 2003. Molecular characterization of Llchit1, a midgut chitinase cDNA from the leishmaniasis vector *Lutzomyia longipalpis*. Insect Biochem. Mol. Biol. 33, 279-287.

Ramasamy, M.S., Raschid, L., Srikrishnaraj, K.A., Ramasamy, R., 1996. Antimidgut antibodies inhibit peritrophic membrane formation in the posterior midgut of *Anopheles tessellatus* (Diptera: Culicidae). J. Med. Entomol. 33, 162-164.

Ramos, A., Mahowald, A., Jacobs-Lorena, M., 1993. Gut-specific genes from the black fly *Simulium vittatum* encoding trypsin-like and carboxypeptidase-like proteins. Insect Mol. Biol. 1, 149-163.

Ramos, A., Mahowald, A., Jacobs-Lorena, M., 1994. Peritrophic matrix of the black fly *Simulium vitattum*: Formation, structure and analysis of its protein components. J. Exp. Zool. 268, 269-281.

Rawlings, N.D., 1998. Introduction: Metallopeptidases and their clans. In: Handbook of Proteolytic Enzymes (ed. by Barrett, A.J., Rawlings, N.D., Woessner, J.F.), Academic Press, New York, 989-991.

Rayms-Keller, A., McGaw, M., Oray, C., Carlson, J.O., Beaty, B.J., 2000. Molecular cloning and characterization of a metal responsive *Aedes aegypti* intestinal mucin cDNA. Insect Mol. Biol. 9, 419-426.

Rebers, J.E., Riddiford, L.M., 1988. Structure and expression of a *Manduca sexta* larval cuticle gene homologous to *Drosophila* cuticle genes. J. Mol. Biol. 203, 411-423.

Rees, D.C., Lewis, M., Lipscomb, W.N., 1983. Refined crystal structure of carboxypeptidase A at 1.54 A resolution. J. Mol. Biol. 168, 367-387.

Rees, J.S., Jarret, P., Ellar, D.J., 2009. Peritrophic membrane contribution to Bt Cry δ-endotoxin susceptibility in Lepidoptera and the effect of calcofluor. J. Invertebr. Pathol. 100, 139-146.

Roelvink, P.W., Corsaro, B.G., Granados, R.R., 1995. Characterization of the *Helicoverpa armigera* and *Pseudaletia unipuncta* granulovirus enhancin genes. J. Gen. Virol. 76, 2693-2705.

Royer, V., Fraichard, S., Bouhin, H., 2002. A novel putative insect chitinase with multiple catalytic domains: Hormonal regulation during metamorphosis. Biochem. J. 366, 921-928.

Ryerse, J.S., Purcell, J.P., Sammons, R.D., Lavrik, P.B., 1992. Peritrophic membrane structure and formation in the larva of a moth, *Heliothis*. Tissue and Cell 24, 751-771.

Sarauer, B.L., 2002. An investigation of the peritrophic membrane proteins from the diamondback moth (*Plutella xylostella*) and the cabbage root maggot (*Delia radicum*). M.Sc. Thesis, University of Saskatchewan, Saskatoon, Canada.

Sarauer, B.L., Gillott, C., Hegedus, D.D., 2003. Characterization of an intestinal mucin from the peritrophic matrix of the diamondback moth, *Plutella xylostella*. Insect Mol. Biol. 12, 333-343.

Schmidt, N.R., Haywood, J.M., Bonning, B.C., 2009. Towards the physiological basis for increased Agrotis ipsilon multiple nucleopolyhedrovirus infection following feeding of *Agrotis ipsilon* larvae on transgenic corn expressing CryFa2. J. Invertebr. Pathol. 102, 141-148.

Schmidt, T., Bleakley, B., Nealson, K.N., 1988. Characterization of an extracellular protease from the insect pathogen *Xenorhabdus luminescens*. Appl. Environ. Microbiol. 54, 2793-2797.

Schorderet, S., Pearson, R.D., Vuocolo, T., Eisemann, C., Riding, G.A., Tellam, R.L., 1998. cDNA and deduced amino acid sequences of a peritrophic membrane glycoprotein, 'Peritrophin-48', from the larvae of *Lucilia cuprina*. Insect Biochem. Mol. Biol. 28, 99-111.

Secundino, N.F.C., Eger-Mangrich, I., Braga, E.M., Santoro, M.M., Pimenta, P.F.P., 2005. *Lutzomyia longipalpis* peritrophic matrix: Formation, structure, and chemical composition. J. Med. Entomol. 42, 928-938.

Semenkovich, C.F., Luo, C., Nakanishi, M.K., Chen, S.H., Smith, L.C., Chan, L., 1990. *In vitro* expression and site-specific mutagenesis of the cloned human lipoprotein lipase gene: Potential *N*-linked glycosylation site asparagine 43 is important for both enzyme activity and secretion. J. Biol. Chem. 265, 5429-5433.

Shahabuddin, M., Lemos, F.J., Kaslow, D.C., Jacobs-Lorena, M., 1996. Antibody-mediated inhibition of *Aedes aegypti* midgut trypsins blocks sporogonic development of *Plasmodium gallinaceum*. Infect. Immun. 64, 739-743.

Shao, L., Devenport, M., Fujioka, H., Ghosh, A., Jacobs-Lorena, M., 2005. Identification and characterization of a novel peritrophic matrix protein, Ae-Aper50, and the microvillar membrane protein, AEG12, from the mosquito, *Aedes aegypti*. Insect Biochem. Mol. Biol. 35, 947-959.

Shao, L., Devenport, M., Jacobs-Lorena, M., 2001. The peritrophic matrix of hematophagous insects. Arch. Insect Biochem. Physiol. 47, 119-125.

Shapiro, M., Dougherty, E.M., 1994. Enhancement in activity of homologous and heterologous viruses against the gypsy moth (Lepidoptera: Lymantriidae) by an optical brightener. J. Econ. Entomol. 87, 361-365.

Shapiro, M., Preisler, H.K., Robertson, J.L., 1987. Enhancement of baculovirus activity on gypsy moth (Lepidoptera: Lymantriidae) by chitinase. J. Econ. Entomol. 80, 1113-1116.

Shen, Z., Jacobs-Lorena, M., 1997. Characterization of a novel gut specific chitinase gene from the human malaria vector, *Anopheles gambiae*. J. Biol. Chem. 272, 28895-28900.

Shen, Z., Jacobs-Lorena, M., 1998. A type I peritrophic matrix protein from the malaria vector *Anopheles gambiae* binds to chitin. Cloning, expression, and characterization. J. Biol. Chem. 273, 17665-17670.

Shen, Z., Jacobs-Lorena, M., 1999. Evolution of chitin-binding proteins in invertebrates. J. Mol. Evol. 48, 341-347.

Shen, Z., Dimopoulos, G., Kafatos, F.C., Jacobs-Lorena, M., 1999. A cell surface mucin specifically expressed in the midgut of the malaria mosquito *Anopheles gambiae*. Proc. Natl. Acad. Sci. USA 96, 5610-5615.

Shi, X., Chamankhah, M., Visal-Shah, S., Hemmingsen, S.M., Erlandson, M., Braun, L. Alting-Mees, M., Khachatourians, G.G., O'grady, M., Hegedus, D.D., 2004. Modeling the structure of the type I peritrophic matrix: Characterization of a *Mamestra configurata* intestinal mucin and novel peritrophin containing 19 chitin binding domains. Insect Biochem. Mol. Biol. 34, 1101-1115.

Simpson, R.M., Newcomb, R.D., Gatehouse, H.S., Crowhurst, R.N., Chagné, D., Gatehouse, L.N., Markwick N.P., Beuning, L.L., Murray, C. Marshall, S.D., Yauk, Y.-K., Nain, B., Wang, Y.-Y., Gleave, A.P., Christeller, J.T., 2007. Expressed sequence tags from the midgut of *Epiphyas postvittana* (Walker) (Lepidoptera: Tortricidae). Insect Mol. Biol. 16, 675-690.

Sivakumar, S., Rajagopal, R., Venkatesh, G.R., Srivastava, A., Bhatnagar, R.K., 2007. Knockdown of aminopeptidase-N from *Helicoverpa armigera* larvae and in transfected Sf21 cells by RNA interference reveals its functional interaction with *Bacillus thuringiensis* insecticidal protein Cry1Ac. J. Biol. Chem. 282, 7312-7319.

Slavicek, J.M., Popham, H.J.R., 2005. The *Lymantria dispar* nucleopolyhedrovirus enhancins are components of occlusion derived virus. J. Virol. 79, 10578-10588.

Smith, G.M., Rothwell, K., Wood, S.L., Yeaman, S.J., Bownes, M., 1994. Specificity and localization of lypolytic activity in adult *Drosophila melanogaster*. Biochem. J. 304, 775-779.

Snodgrass, R.E., 1956. Anatomy of the honey bee. Comstock Publishing Associates, Ithaca, USA.

Spence, K.D., Kawata, M.Y., 1993. Permeability characteristics of the peritrophic membranes of *Manduca sexta* larvae. J. Insect Physiol. 39, 785-790.

Stamm, B., D'Haese, J., Peters, W., 1978. SDS gel electrophoresis of proteins and glycoproteins from peritrophic membranes of some Diptera. J. Insect Physiol. 24, 1-8.

Statistica, 1997. Statistica release 5.1. Tulsa, USA.

Stöcker, W., Gomis-Rüth, F.-X., Bode, W., Zwilling, R., 1993. Implications of the three-dimensional structure of astacin for the structure and function of the astacin family of zinc-endopeptidases. Eur. J. Biochem. 214, 215-231.

Strobl, S., Maskos, K., Betz, M., Wiegand, G., Huber, R., Gomis-Rüth, F-X., Glockshuber, R., 1998. Crystal structure of yellow mealworm alpha-amylase at 1.64 A resolution. J. Mol. Biol. 278, 617-628.

Sudha, P.M., Muthu, S.P., 1988. Damage to the midgut epithelium caused by food in the absence of peritrophic membrane. Curr. Sci. 57, 624-625.

Suetake, T., Tsuda, S., Kawabata, S., Miura, K., Iwanaga, S., Hikichi, K., Nitta, K., Kawano, K., 2000. Chitin-binding proteins in invertebrates and plants comprise a common chitin binding structural motif. J. Biol. Chem. 275, 17929-17932.

Summers, C.B., Felton, G.W., 1996. Peritrophic envelope as a functional antioxidant. Arch. Insect Biochem. Physiol. 32, 131-142.

Sutherland, G.B., Ewen, A.B., 1974. Fecundity decrease in mosquitoes ingesting blood from specifically sensitized mammals. J. Insect Physiol. 20, 655-660.

Takase, H., Watanabe, A., Yoshizawa, Y., Kitami, M., Ryoichi, S., 2009. Identification and comparative analysis of three novel c-type lectins from the silkworm with functional implications in pathogen recognition. Dev. Comp. Immunol. 33, 789-800.

Tanada, Y., 1959. Synergism between two viruses of the armyworm *Pseudaletia unipuncta* (Haworth) (Lepidoptera, Noctuidae) J. Insect Pathol. 1, 215-231.

Tanji, T., Ohaski-Kobayashi, A., Natori, S., 2006. Participation of a galactose-specific c-type lectin in *Drosophila* immunity. Biochem. J. 396, 127-138.

Tellam, R.L., 1996. The peritrophic matrix. In: The Biology of the Insect Midgut (ed. by Billingsley, P.F., Lehane, M.J.), Chapman and Hall, London, 86-114.

Tellam, R.L., Eisemann, C., 2000. Chitin is only a minor component of the peritrophic matrix from larvae of *Lucilia cuprina.* Insect Biochem. Mol. Biol. 30, 1189-2001.

Tellam, R.L., Eisemann, C., Casu, R., Pearson, R., 2000. The intrinsic peritrophic matrix protein peritrophin-95 from larvae of *Lucilia cuprina* is synthesized in the cardia and regurgitated or excreted as a highly immunogenic protein. Insect Biochem. Mol. Biol. 30, 9-17.

Tellam, R.L., Vuocolo, T., Eisemann, C., Briscoe, S., Elvin, C., Pearson, R., 2003. Identification of an immuno-protective mucin-like protein, peritrophin-55, from the peritrophic matrix of *Lucilia cuprina* larvae. Insect Biochem. Mol. Biol. 33, 239-252.

Tellam, R.L., Wijffels, G., Willadsen, P., 1999. Peritrophic matrix proteins. Insect Biochem. Mol. Biol. 29, 87-101.

Terenius, O., Papanicolaou, A., Garbutt, J.S., Eleftherianos, I., Huvenne, H., Sriramana, K., Albrechtsen, M., An, C., Aymeric, J-L., Barthel, A., Bebas, P., Bitra, K., Bravo, A., Chevalier, F., Collinge, D.P., Crava, C.M., de Maagd, R.A., Duvic, B., Erlandson, M., Faye, I., Felföldi, G., Fujiwara, H., Futahashi, R., Gandhe, A.S., Gatehouse, H.S., Gatehouse, L.N., Giebultowicz, J., Gómez, I., Grimmelikhuijzen, C.J., Groot, A.T., Hauser, F., Heckel, D.G., Hegedus, D.D., Hrycaj, S., Huang, L., Hull, J.J., Iatrou, K., Iga, M., Kanost, M.R., Kotwica, J., Li, C., Li, J., Liu, J., Lundmark, M., Matsumoto, S., Meyering-Vos, M., Millichap, P.J., Monteiro, A., Mrinal, N., Niimi, T., Nowara, D., Ohnishi, A., Oostra, V., Ozaki, K., Papakonstantinou, M., Popadic, A., Rajam, M.V., Saenko, S., Simpson, R.M., Soberón, M., Strand, M.R., Tomita, S., Toprak, U., Wang, P., Wee, C.W., Whyard, S., Zhang, W., Nagaraju, J., ffrench-Constant, R.H., Herrero, S., Gordon, K., Swevers, L., Smagghe, G., 2011. RNA interference in Lepidoptera: An overview of successful and unsuccessful studies and implications for experimental design. J. Insect Physiol. 57, 231-245.

Terra, W.R., 2001. The origin and functions of the insect peritrophic membrane and peritrophic gel. Arch. Insect Biochem. Physiol. 47, 47-61.

Terra, W.R., Ferreira, C., 1981. The physiological role of the peritrophic membrane and trehalase: Digestive enzymes in the midgut and excreta of starved larvae of *Rhynchosciara*. J. Insect Physiol. 27, 325-331.

Terra, W.R., Ferreira, C., 1994. Insect digestive enzymes: Properties, compartmentalization and function. Comp. Biochem. Physiol. 109B, 1-62.

Terra, W.R., Ferreira, C., de Bianchi, A.G., 1979. Distribution of digestive enzymes among the endo- and ectoperitrophic spaces and midgut cells of *Rhynchosciara* and its physiological significance. J. Insect Physiol. 25, 487-494.

Thiem, S.M., 2009. Baculovirus genes affecting host function. In Vitro Cell. Dev. Biol.-Animal 45, 111-126.

Titani, K., Ericsson, L.H., Walsh, K.A., Neurath, H.I., 1975. Amino-acid sequence of bovine carboxypeptidase B. Proc. Natl. Acad. Sci. USA 72, 1666-1670.

Tomoyasu, Y., Denell, R.E., 2004. Larval RNAi in *Tribolium* (Coleoptera) for analyzing adult development. Dev. Genes. Evol. 214, 575-578.

Tomoyasu, Y., Miller, S.C., Tomita, S., Schoppmeier, M., Grossmann, D., Bucher, G., 2008. Exploring systemic RNA interference in insects: A genome-wide survey for RNAi genes in *Tribolium*. Genome Biol. 9, R10.

Towbin, H., Staehelin, T., Gordon, J., 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. Proc. Natl. Acad. Sci. USA 76, 4350-4354.

Trudel, J., Asselin, A., 1989. Detection of chitinase activity after polyacrylamide gel electrophoresis. Anal. Biochem. 178, 362-366.

Trudel, J., Asselin, A., 1990. Detection of chitin deacetylase activity after polyacrylamide gel electrophoresis. Anal. Biochem. 189, 249-253.

Tsigos, I., Martinou, A., Kafetzopoulos, D., Bouriotis, V., 2000. Chitin deacetylases: New, versatile tools in biotechnology. Trends Biotechnol. 18, 305-312.

Turner, C.T., Davy, M.W., MacDiarmid, R.M., Plummer, K.M., Birch, N.P., Newcomb, R.D., 2006. RNA interference in the light brown apple moth, *Epiphyas postvittana* (Walker) induced by double-stranded RNA feeding. Insect Mol. Biol. 15, 383-391.

Turnock, W.J., 1985. Developmental, survival, and reproductive parameters of bertha armyworm, *Mamestra configurata* (Lepidoptera: Noctuidae) on four plant species. Can. Entomol. 117, 1267-1271.

Van Klinken, B.J.-W., Dekker, J., Buller, H.A., Einerhand, A.W.C., 1995. Mucin gene structure and expression: Protection vs. adhesion. Am. J. Physiol. 269G, 613-627.

Van Scheltinga, A.C.T., Kalk, K.H., Beintema, J.J., Dijkstra, B.W., 1994. Crystal structures of hevamine, a plant defence protein with chitinase and lysozyme activity, and its complex with an inhibitor. Structure 2, 1181-1189.

Venancio, T.M., Cristofoletti, P.T., Ferreira, C., Verjovski-Almeida, S., Terra, W.R., 2009. The *Aedes aegypti* larval transcriptome: A comparative perspective with emphasis on trypsins and the domain structure of peritrophins. Insect Mol. Biol. 18, 33-44.

Vogt, G., Stöcker, W., Storch, V., Zwilling, R., 1989. Biosynthesis of *Astacus* protease, a digestive enzyme from crayfish. Histochemistry 91, 373-381.

Wagner, W., Möhrlen, F., Schnetter, W., 2002. Characterization of the proteolytic enzymes in the midgut of the European cockchafer, *Melolontha melolontha* (Coleoptera: Scarabaeidae). Insect Biochem. Mol. Biol. 32, 803-814.

Walker, V.K., Geer, B.W., Williamson, J.H., 1980. Dietary modulation and histochemical localization of leucine aminopeptidase activity in *Drosophila melanogaster* larvae. Insect Biochem. 10, 543-548.

Wang, E.C., Hung, S.H., Cahoon, M., Hedstrom, L., 1997. The role of the Cys191-Cys220 disulphide bond in trypsin: New targets for engineering substrate specificity. Protein Eng. 10, 405-411.

Wang, J-L., Jiang, X-J., Wang, Q., Hou, L-J., Xu, D-W., Wang, J-X., Zhao, X-F., 2007. Identification and expression profile of a putative basement membrane protein gene in the midgut of *Helicoverpa armigera*. BMC Dev. Biol. 7, 76.

Wang, P., Granados, R.R., 1997a. An intestinal mucin is the target substrate for a baculovirus enhancin. Proc. Natl. Acad. Sci. USA 94, 6977-6982.

Wang, P., Granados, R.R., 1997b. Molecular cloning and sequencing of a novel invertebrate intestinal mucin. J. Biol. Chem. 272, 16663-16669.

Wang, P., Granados, R.R., 1998. Observations on the presence of the peritrophic membrane in larval *Trichoplusia ni* and its role in limiting baculovirus infection. J. Invertebr. Pathol. 72, 57-62.

Wang, P., Granados, R.R., 2000. Calcofluor disrupts the midgut defence system in insects. Insect Biochem. Mol. Biol. 30, 135-143.

Wang, P., Granados, R.R., 2001. Molecular structure of the peritrophic membrane (PM): Identification of potential PM target sites for insect control. Arch. Insect Biochem. Physiol. 47, 110-118.

Wang, P., Hammer, D.A., Granados, R.R., 1994. Interaction of *Trichoplusia ni* granulosis virus-encoded enhancin with the midgut epithelium and peritrophic membrane of four lepidopteran insects. J. Gen. Virol. 75, 1961-1967.

Wang, P., Li, G., Granados, R.R., 2004a. Identification of two new peritrophic membrane proteins from larval *Trichoplusia ni*: Structural characteristics and their functions in the protease rich insect gut. Insect Biochem. Mol. Biol. 34, 215-227.

Wang, P., Li, G., Kain, W., 2004b. Characterization and cDNA cloning of midgut carboxypeptidases from *Trichoplusia ni*. Insect Biochem. Mol. Biol. 34, 831-843.

Wang, S., Jayaram, A.S., Hemphala, J., Senti, K.A., Tsarouhas, V., Jin, H., Samakovlis, C., 2006. Septate-junction-dependent luminal deposition of chitin deacetylases restricts tube elongation in the *Drosophila* tracheae. Curr. Biol. 16, 180-185.

Wang, X., Ding, X., Gopalakrishnan, B., Morgan, T.D., Johnson, L., White, F., Muthukrishnan, S., Kramer, K.J., 1996. Characterization of a 46 kDa insect chitinase from transgenic tobacco. Insect Biochem. Mol. Biol. 26, 1055-1064.

Watanabe, T., Kobori, K., Kiyotaka, M., Fujii, T., Sakai, H., Uchida, M., Tanaka, H., 1993. Identification of glutamic acid 204 and aspartic acid 200 in chitinase A1 of *Bacillus circulans* WL-12 as essential residues for chitinase activity. J. Biol. Chem. 268, 18567-18572.

Waterhouse, D.F., 1953a. Occurrence and endodermal origin of the peritrophic membrane in some insects. Nature 172, 676-677.

Waterhouse, D.F., 1953b. The occurrence and significance of the peritrophic membrane, with special reference to adult Lepidoptera and Diptera. Aust. J. Zool. 1, 299-318.

Waterhouse, D.F., 1954. The rate of production of the peritrophic membrane in some insects. Aust. J. Biol. Sci. 7, 59-72.

Waterhouse, D.F., 1957. Digestion in insects. Annu. Rev. Entomol. 2, 1-18.

Waters, L.C., Zelhof, A.C., Shaw, B.J., Chang, L.Y., 1992. Possible involvement of the long terminal repeat of transposable element 17.6 in regulating expression of an insecticide resistance-associated P450 gene in *Drosophila*. Proc. Natl. Acad. Sci. USA 89, 4855-4859.

Weir, J.S., 1957. The functional anatomy of the midgut of larvae of the ant, *Myrmica*. Q. J. Microsc Sci. S3-98, 499-506.

Wigglesworth, V.B., 1929. Digestion in the tsetse-fly: A study of structure and function. Parasitology 21, 288-321.

Wigglesworth, V.B., 1930. The formation of the peritrophic membrane in insects, with special reference to the larvae of mosquitoes. Q. J. Microsc. Sci. 73, 583-616.

Wigglesworth, V.B., 1972. The Principles of Insect Physiology. Seventh edition, Chapman and Hall, London.

Wijffels, G., Eisemann, C., Riding, G., Pearson, R., Jones, A., Willadsen, P., Tellam, R., 2001. A novel family of chitin-binding proteins from insect type 2 peritrophic matrix: cDNA sequences, chitin binding activity, and cellular localization. J. Biol. Chem. 276, 15527-15536.

Wong, C., Sridhara, S., Bardwell, J.C.A., Jakob, U., 2000. Heating greatly speeds Coomassie blue staining and destaining. Biotechniques 28, 426-432.

Xu, G.Y., Ong, E., Gilkes, N.R., Kilburn, D.G., Muhandiram, D.R., Harris-Brandts, M., Carver, J.P., Kay, L.E., Harvey, T.S., 1995. Solution structure of a cellulose binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. Biochemistry 34, 6993-7009.

Yamamoto, T., Tanada, Y., 1978. Phospholipid, an enhancing component in the synergistic factor of a granulosis virus of the armyworm, *Pseudaletia unipuncta*. J. Invertebr. Pathol. 31, 48-56.

Yan, J., Cheng, Q., Li, C.B., Aksoy, S., 2002. Molecular characterization of three gut genes from *Glossina morsitans morsitans*: cathepsin B, zincmetalloprotease and zinc-carboxypeptidase. Insect Mol. Biol. 11, 57-65.

Yang, Y-J., Davies, D.M., 1968. Digestion, emphasizing trypsin activity in simuliids (Diptera) fed blood, blood-scurose mixtures and sucrose. J. Insect Physiol. 14, 205-222.

Yang, Y., Zhu, Y.C., Ottea, J., Husseneder, C., Leonard, B.R., Abel, C., Huang, F., 2010. Molecular characterization and RNA interference of three midgut aminopeptidase N isozymes from *Bacillus thuringiensis*-susceptible and -resistant strains of sugarcane borer, *Diatraea saccharalis*. Insect Biochem. Mol. Biol. 40, 592-603.

Ye, S., Goldsmith, E.J., 2001. Serpins and other covalent protease inhibitors. Curr. Opin. Struct. Biol. 11, 740-745.

Yin, J., Wei, Z-J., Li, K-B., Cao, Y-Z., Guo, W., 2010. Identification and molecular characterization of a new member of the peritrophic membrane proteins from the meadow moth, *Loxostege sticticalis*. Int. J. Biol. Sci. 6, 491-498.

Yu, X.-Q., Zhu, Y.-F., Ma, C., Fabrick, J.A., Kanost, M.R., 2002. Pattern recognition proteins in *Manduca sexta* plasma. Insect Biochem. Mol. Biol. 32, 1287-1293.

Zhang, P., Aso, Y., Yamamoto, K., Banno, Y., Wang, Y., Tsuchida, K., Kawaguchi, Y., Fujii, H., 2006. Proteome analysis of silk gland proteins from the silkworm, *Bombyx mori*. Proteomics 6, 2586-2599.

Zheng, Y.P., Retnakaran, A., Krell, P.J., Arif, B.M., Primavera, M., Feng, Q.L., 2003. Temporal, spatial and induced expression of chitinase in the spruce budworm, *Choristoneura fumiferana*. J. Insect Physiol. 49, 241-247.

Zhou, H-X., Tan, X-M., Li, C-Y., Wang, J-P., Sun, X-G., Guo, W., Li, G-X., 2009. Molecular cloning and sequencing of cDNAs of two peritrophic membrane proteins from *Holotrichia oblita* (Coleoptera: Melolonthidae). Acta Entomol. Sinica 52, 10-16.

Zhu, Q., Arakane, Y., Banerjee, D., Beeman, R.W., Kramer, K.J., Muthukrishnan, S., 2008a. Domain organization and phylogenetic analysis of the chitinase-like family of proteins in three species of insects. Insect Biochem. Mol. Biol. 38, 452-466.

Zhu, Q., Arakane, Y., Beeman, R.W., Kramer, K.J., Muthukrishnan, S., 2008b. Functional specialization among insect chitinase family genes revealed by RNA interference. Proc. Natl. Acad. Sci. USA 105, 6650-6655.

Zhu, R., Liu, K., Peng, J., Yang, H., Hong, H., 2007. Optical brightener M2R destroys the peritrophic membrane of *Spodoptera exigua* (Lepidoptera: Noctuidae) larvae. Pest Manag. Sci. 63, 296-300.

Zhu, X., Zhang, H., Fukamizo, T., Muthukrishnan, S., Kramer, K.J., 2001. Properties of *Manduca sexta* chitinase and its C-terminal deletions. Insect Biochem. Mol. Biol. 31 1221-1230.

Zhu, Y., Hukuhara, T., Tamura, K., 1989. Location of a synergistic factor in the capsule of a granulosis virus of the armyworm, *Pseudaletia unipuncta.* J. Invertebr. Pathol. 54, 49-56.

Zhuzhikov, D.P., 2003. Formation and destruction of the peritrophic envelope in the cockroach *Nauphoeta cinerea* (Blattodea: Blaberidae). Zoologičeskij Žurnal 2, 344-353.

Zimoch, L., Merzendorfer, H., 2002. Immunolocalization of chitin synthase in the tobacco hornworm. Cell Tissue Res. 308, 287-297.

Zimoch, L., Hogenkamp, D., Kramer, K.J., Muthukrishnan, S., Merzendorfer, H., 2005. Regulation of chitin synthesis in the larval midgut of *Manduca sexta*. Insect Biochem. Mol. Biol. 35, 515-527.

APPENDICES

Appendix A. Data from *M. configurata* peritrophic matrix proteins from LC-MS/MS spectra.

Protein	Accession	#UP	SC (%)	MS	Peptide Sequence	m/z	z	Delta	Modification
1. PERITROPHINS									
1.1 Non-mucin peritrophins									
McPM1	AY277403	3	10	1929	R.IVPDDCACNPR.N	658.766	2	-0.0475	none
					R.CDWPSNVVCGDR.I	732.8184	2	0.0301	none
					R.CDWPEHVQCGDR.V	779.7867	2	-0.0501	none
McCBD3P	HM357863	1	4	61	K.ILLELEDGETK.F	630.4402	2	0.2003	none
McPPAD1	GU596430	1	10	47	R.VFENVDPNDLSCDPAGHIFLLLPHFTDCSK.F	1153.2156	3	0.0131	none
1.2 Insect intestinal mucins									
McIIM1	AY057052	13	17	3214	R.NIPQITR.H	421.3268	2	0.1574	none
					-.LIYLTIIK.-	488.7642	2	-0.1230	none
					K.YYQCDAGK.K	502.7179	2	0.0143	none
					K.YYVCDNGR.L	523.7270	2	0.0107	none
					K.YYTCVGDEFR.V	655.2223	2	-0.1145	none
					R.DVIDLIEHELNAEL.-	811.9470	2	0.0595	none
					K.EECNVAPWAHAECDK.Y	605.9174	3	-0.0048	none
					R.HVEGMLMFIPHDFNNR.D	652.9719	3	-0.0196	none
					R.HVEGMLMFIPHDFNNR.D	658.3287	3	0.0560	Oxi. (M)
					R.HVEGMLMFIPHDFNNR.D	663.6350	3	-0.0201	2 Oxi. (M)
					K.CKEECNVAPWAHAECDK.Y	701.9544	3	-0.0195	none
					K.LIPHEEYCHLFYYCDK.G	729.6553	3	-0.0161	none
					R.VNACAEGLHFNPSTLTCDFICNAGCVR.N	1028.4842	3	0.0790	none
McIIM4	FJ670569	5	15	738	K.YYACDNGR.L	509.6939	2	-0.0242	none
					K.FYQCNFGQK.V	596.2695	2	0.0065	none
					K.GELVLSSCPEPLYFDPK.A	976.0443	2	0.1296	none
					K.LVPHEEYCHLFYYCDK.G	724.9696	3	-0.0575	none
					R.DCAPGTVFHFEIQVCDWPR.N	778.7038	3	0.0538	none
McIIM2	FJ670567	4	17	605	R.HNVDCTDR.Y	508.7213	2	0.0139	none
					R.QCPATLHFNR.V	622.2707	2	-0.0659	none
					K.CPEEQETDWTIELLLR.H	1016.5183	2	0.0602	none
					K.CTFGKPIEYTCPADLWFDLDTQQCDWR.H	1141.5352	3	0.0995	none
2. ENZYMES									
2.1 Chitin deacetylases									
McCDA2	HM357864	12	41	4496	K.QEIADQK.G	416.2148	2	0.0017	none
					K.NPIPVGEYR.E	522.7672	2	-0.0202	none
					R.VYPWTGNPLGQ.-	616.3107	2	0.0035	none
					R.CSSTDIPGGLLPR.D	686.8273	2	-0.0417	none
					K.GQISHFANIPFESIK.G	844.4418	2	-0.0040	none
					R.IPQDYWATASYDVIK.Q	885.4408	2	-0.0002	none
					R.DTPQFVTITFDDGVNVR.N	962.4846	2	0.0171	none
					R.FVNELYNQGYEIALHSISHR.I	797.4002	3	-0.0028	none
					R.FTDLISNLSDAFMVNASEVIDWVK.N	905.4727	3	0.0654	none
					R.FTDLISNLSDAFMVNASEVIDWVK.N	910.8003	3	0.0534	Oxi. (M)
					R.APFGFYVHEWYLASNPAVQAAFIR.F	918.8098	3	0.0321	none
					R.EWSTASCAAQPDCGPLLGPNAMNYWMGACNVCPR.V	1281.2080	3	-0.0012	none
McCDA1	EU660852	13	38	2224	R.ADEDAWYK.F	499.2398	2	0.0462	none
					K.FILTNFER.H	520.3736	2	0.1827	none
					R.VYPWLGNPLGQ.-	622.3354	2	0.0166	none
					R.CSSTNIPGGLQPR.D	693.7936	2	-0.0995	none
					K.IQIAHFANIPYESIK.G	872.4770	2	0.0040	none
					K.EAEECTPETVCELPNCR.C	698.6203	3	-0.0110	none
					R.APFGFFVHEAFLSAFPAVR.G	704.0787	3	0.1307	none
					R.DTPQFVTVTFDDGINVNNILTYR.N	881.4476	3	0.0186	none
					R.SCGPLNSGHNGMDYWMQICNVCPR.V	951.7513	3	0.0612	none
					R.FLDLINNLPDTFMVNSHEVIDWVK.N	953.8783	3	0.1817	none
					R.SCGPLNSGHNGMDYWMQICNVCPR.V	957.0760	3	0.0403	Oxi. (M)

Appendix A. (continued)

Protein	Accession	#UP	SC (%)	MS	Peptide Sequence	m/z	z	Delta	Modification
					R.FLDLINNLPDTFMVNSHEVIDWVK.N	959.2159	3	0.1997	Oxi. (M)
					R.SCGPLNSGHNGMDYWMQICNVCPR.V	962.4112	3	0.0510	2 Oxi. (M)
2.2 Putative digestive enzymes									
2.2.1 Serine proteases									
2.2.1.1. Trypsins									
McSP1	FJ205402	3	13	500	R.IVGGSVTTIDR.Y	559.3057	2	-0.0171	none
					R.SASTFSFNNNVR.A	672.3969	2	0.1526	none
					R.AILTAAHCTVGDAVGR.W	538.0073	3	0.1801	none
McSP57	HM990179	4	22	379	R.IVGGSLTTIDR.Y	566.3138	2	-0.0165	none
					R.SAGNFVFNNNVR.A	669.9092	2	0.1562	none
					R.IAHPTFNLNTLAGDIAILR.S	684.0557	3	0.0082	none
					R.DQCTYDEGGPLYHNGVVVGVR.S	779.1185	3	0.2637	none
McSP52	HM990174	5	17	322	R.SATTIAFNNNVR.A	654.3339	2	-0.0097	none
					R.HVQIYSINQATCR.S	530.6645	3	0.1937	none
					R.YPSLTDNMLCSGVLNVGGR.D	685.1055	3	0.3177	none
					R.YPSLTDNMLCSGVLNVGGR.D	690.4567	3	0.3764	Oxi. (M)
					R.SRYPSLTDNMLCSGVLNVGGR.D	771.4934	3	0.3533	Oxi. (M)
McSP34	FJ205445	2	11	125	R.VSSFSNWISANQ.-	670.3352	2	0.0354	none
					R.SASSIVYNNNVQR.A	726.4671	2	0.2033	none
McSP62	HM990184	1	5	69	K.LSSPAVFSDR.I	539.8441	2	0.1281	none
McSP60	HM990182	1	2	60	R.LDNPAVFSSR.V	553.3671	2	0.1633	none
McSP2	FJ205440	1	6	41	R.YVLSAAHCFEGALFQINYR.R	753.7460	3	0.1219	none
McSP3	FJ205433	1	2	38	R.YASLTIPQR.V	524.8534	2	0.1210	none
2.2.1.2 Chymotrypsins									
McSP24	FJ205398	11	30	3752	R.AISNESCGR.T	497.2223	2	-0.0045	none
					K.VPRPFVYSR.L	560.807	2	-0.0195	none
					R.TWTLGSVTEVLR.W	681.4117	2	0.0738	none
					K.EWVITAAHCTAGR.L	736.3577	2	-0.0029	none
					R.DYDGEQVYASGHGR.T	777.3271	2	-0.0146	none
					R.TFVSGIIIDTSICAR.F	551.6271	2	-0.0008	none
					R.MVSPTGGVSGCGGSIVSK.E	840.3967	2	-0.0231	none
					R.MVSPTGGVSGCGGSIVSK.E	848.3759	2	-0.0596	Oxi. (M)
					K.YNNDLPLLVQPHDISLVK.V	693.3722	3	-0.0259	none
					R.IVSGWEAQPGQHPHHAALR.M	697.6877	3	-0.0146	none
					R.GGVVSLTDPEYIFETTEYYNHPK.Y	887.0899	3	-0.0010	none
McSP27	FJ205413	13	32	2631	R.FTVVLGSVR.L	489.3857	2	0.1863	none
					R.RFTVVLGSVR.L	567.3283	2	-0.0296	none
					R.ILIGVTSFGSDR.G	632.8006	2	-0.0956	none
					R.GCQVGFPAAFAR.V	640.9364	2	0.2451	none
					R.VTSYISWINQR.L	683.9642	2	0.2098	none
					R.VLTAAHCWFDGR.N	478.2169	3	-0.0428	none
					R.STCQGDSGGPLVVTSNNR.R	924.9056	2	-0.0466	none
					R.LNTASVAMHGSWNPNLVR.N	656.3273	3	-0.0241	none
					R.LNTASVAMHGSWNPNLVR.N	661.6600	3	-0.0211	Oxi. (M)
					R.STCQGDSGGPLVVTSNNRR.I	668.9716	3	-0.0514	none
					R.QSFAIIQSSNICTSGANGR.S	671.0643	3	0.2121	none
					R.DGGSVSGALSHVNLPVITNAVCR.Q	775.0664	3	0.0022	none
					R.IVGGSTSNLGQFPYQAGLLADFSGGQGVCGGSLLNSR.R	1228.9557	3	0.0450	none
McSP51	HM990173	8	43	1223	R.FTVVLGSVR.L	489.3857	2	0.1863	none
					R.ILIGVTSFGSDR.G	632.8006	2	-0.0956	none
					R.GCQVGFPAAFAR.V	640.9364	2	0.2451	none
					R.VTSFISWINQR.L	675.9420	2	0.1602	none
					R.VLTAAHCWFDGR.N	478.2169	3	-0.0428	none
					R.STCQGDSGGPLVVTSNNR.R	924.9056	2	-0.0466	none
					R.STCQGDSGGPLVVTSNNRR.I	668.9716	3	-0.0514	none
					R.DGGSVSGSLSHVNLPVITNAVCR.Q	780.3803	3	-0.0508	none
McSP61	HM990183	9	37	1145	R.DGGSVSETLR.H	510.8128	2	0.1227	none
					R.SVEVVLGSVR.L	522.8616	2	0.1111	none

Appendix A. (continued)

Protein	Accession	#UP	SC (%)	MS	Peptide Sequence	m/z	z	Delta	Modification
					R.GCAVGSPAVFAR.V	596.2851	2	-0.0310	none
					R.ILVGVTSFGSGR.G	596.9492	2	0.2227	none
					R.GDSGGPLTVNSNGR.R	665.8969	2	0.1520	none
					R.VTSFISWINQR.L	675.9420	2	0.1602	none
					R.VLTAAHCWFDGR.N	478.2169	3	-0.0428	none
					R.HVNLPVITNAVCR.Q	498.2565	3	-0.0503	none
					R.IVGGSNANPGQFPFQAGLLIDVAAGR.S	857.2222	3	0.2999	none
McSP23	FJ205424	10	31	890	R.NDVAMIR.L	409.7530	2	0.0799	none
					R.SFTVVLGSTR.L	533.7833	2	-0.0298	none
					R.GCQVGSPAAFAR.V	610.8658	2	0.1403	none
					R.STCNGDSGGPLTVNR.S	767.8349	2	-0.0290	none
					R.SGKPILIGITSFGSAR.G	535.3033	3	-0.0214	none
					R.VITAAHCWFDGQNQAR.S	625.3065	3	0.0287	none
					R.VSTSNVVMHGSWNPSWIR.N	686.4153	3	0.2293	none
					R.SSFPLYVTDSNICVSGAGGR.S	1043.9707	2	-0.0521	none
					R.TADGSAGAITTNQGLNHVTVPVITNAVCR.S	979.8107	3	-0.0672	none
					R.IVGGSTSNLGQFPYQAGLLADFSGGQGVCGGSLLNSR.R	1228.9557	3	0.0450	none
McSP28	FJ205404	2	13	888	R.AILIGITSFGSAR.G	653.3641	2	-0.0316	none
					R.TADSSAGAITTSQGLR.H	768.3840	2	-0.0053	none
McSP55	HM990177	8	34	759	R.SFTVVLGSIR.L	539.8115	2	-0.0098	none
					R.NDVAMINLPR.A	571.7917	2	-0.0226	none
					R.NDVAMINLPR.A	579.8000	2	-0.0009	Oxi. (M)
					R.GCQTGAPDAFAR.V	625.7668	2	-0.0320	none
					R.ILVGVTSFGSMR.G	641.8305	2	-0.0287	Oxi. (M)
					R.VTSYISWINQR.L	683.9642	2	0.2098	none
					R.DGASVSGALSHVNLPVITNAVCR.Q	779.7316	3	-0.0177	none
					R.AVGNTANISPIALPSGNELNNQFNGATATASGFGR.T	1144.5424	3	-0.0812	none
McSP30	FJ205436	8	20	606	R.CGGALVSLR.H	466.7722	2	0.0389	none
					K.FTIHPNYR.T	524.2836	2	0.0230	none
					R.IIGGTLASDTQFR.Y	689.8702	2	0.0006	none
					R.YMVSLQQLSQVTNYMR.G	654.3574	3	0.0956	none
					R.YMVSLQQLSQVTNYMR.G	659.6869	3	0.0894	Oxi. (M)
					R.YMVSLQQLSQVTNYMR.G	659.6947	3	0.1129	none
					R.YMVSLQQLSQVTNYMR.G	997.0219	3	0.0847	2 Oxi. (M)
					R.HAVTIASCLHGANGAVIDPAQYR.V	807.7783	3	0.1112	none
McSP56	HM990178	5	18	372	R.SLTVILGSVR.L	522.8099	2	-0.0285	none
					R.DNGAVSETLR.H	531.2496	2	-0.0303	none
					R.VTSFISWINQR.L	675.9420	2	0.1602	none
					R.HVNLPVITNAVCR.Q	498.2565	3	-0.0503	none
					R.VLTAAHCWFDGQNQAR.S	625.3065	3	0.0287	none
McSP48	FJ205437	5	28	336	R.GTSLSSAALVSPTR.V	673.8446	2	-0.0456	none
					K.NICTGDSGGPLAVTR.K	759.3657	2	-0.0136	none
					R.VLTAAHNLNDAVVSVTK.V	584.6317	3	-0.0845	none
					R.IVTTDFVLHENWTPAIIR.N	709.0451	3	-0.0232	none
					R.IIGGSLAYLGQFPYQVGLLTENPR.G	869.4663	3	-0.0133	none
McSP25	FJ205416	4	20	329	K.YIHPEYDEIR.A	445.5421	3	-0.0256	none
					K.NFNFAGVPMIVSGFGR.T	865.0280	2	0.1961	Oxi. (M)
					R.TDDLWNGGAASEILLWTLQR.G	753.7570	3	0.1162	none
					R.AGVQTADLALVGLEQEISYSNNVQPSR.L	953.8171	3	-0.0115	none
McSP29	FJ205412	6	31	300	R.NDVAMIR.L	409.7530	2	0.0799	none
					R.SFTVVLGSNR.L	540.3534	2	0.1152	none
					R.GCQIGSPAAFAR.V	617.8029	2	-0.0011	none
					R.STCQGDSGGPLVVNR.S	773.8668	2	-0.0015	none
					R.VVTAAHCWFDGQNQAR.S	620.6352	3	0.0306	none
					R.NSFPLIIQNSNICTSGAGGR.S	702.6689	3	-0.0476	none
McSP58	HM990180	3	13	263	R.ELVFGDTIQPIR.I	694.3803	2	-0.0048	none
					R.IVSGWEAQEGQFPHQLSLR.M	728.0371	3	-0.0073	none

Appendix A. (continued)

Protein	Accession	#UP	SC (%)	MS	Peptide Sequence	m/z	z	Delta	Modification
					K.YYNHPLYNEALQSVVQPNDIGLIEFGR.E	1050.5366	3	0.0262	none
McSP53	HM990175	4	14	250	R.VTSFMDFFR.Q	575.3394	2	0.1318	none
					R.VTSFMDFFR.Q	583.3403	2	0.1386	Oxi. (M)
					K.VITAAHCWFDGR.N	478.2169	3	-0.0428	none
					R.FAFPLNLHASNICTSSLGGASVCR.G	860.4032	3	-0.0542	none
McSP32	FJ205441	5	17	209	K.TIYSLDR.C	434.2444	2	0.0245	none
					K.LSEIQSTTR.G	517.7731	2	-0.0088	none
					R.FPENAVTIVR.L	573.3454	2	0.0523	none
					K.NCIASENPEVYSR.V	769.8749	2	0.0522	none
					R.ISNALPAAENQFPYAISLQK.L	725.7446	3	0.0748	none
McSP54	HM990176	3	16	203	K.ISELQSYTR.G	548.8050	2	0.0394	none
					R.VFAGAVTLTSDNSPYHIR.N	650.0266	3	0.0728	none
					R.IVEGATASVNQFQYMVSLQK.I	743.7405	3	0.0850	Oxi. (M)
2.2.1.3 Elastases									
McSP38	FJ205426	6	39	5120	K.DEDGALSVSSDEVEVTR.Y	904.3666	2	-0.0932	none
					R.YGVVNVINPSLVTENSLVR.I	691.8447	3	0.3856	none
					R.LVESTENIEPIGIGESSDAGK.F	1073.0135	2	-0.0360	none
					R.IHPDYNLASGENNIALININR.L	784.4178	3	0.0288	none
					K.FCAFGGPDGPGEILNCYDVEVSK.D	844.3755	3	-0.0097	none
					R.YDVGAALVSDGAQVGVLVDDSGNFIATAK.Y	951.4743	3	-0.0228	none
McSP35	FJ205405	5	18	505	R.FTVVLGSIR.L	496.3554	2	0.1101	none
					R.GCAIGDPAAYAR.V	611.3746	2	0.1739	none
					R.NILIGVTSFGTGR.G	667.9501	2	0.1502	none
					R.VTSYISWINQR.L	683.9642	2	0.2098	none
					R.VASSNVDMHANWTPSLIR.N	672.0942	3	0.2871	Oxi. (M)
McSP17	FJ205400	4	12	423	R.VTSYVSFFNQHL.-	721.3547	2	-0.0090	none
					R.VTVVLGSTTLFSGGTR.L	798.0620	2	0.2368	none
					R.LDSSVIAMHPNWTPALIR.N	674.4659	3	0.3194	none
					R.LDSSVIAMHPNWTPALIR.N	679.6838	3	-0.0218	Oxi. (M)
McSP65	HM990187	4	13	390	R.VTVVLGSIR.L	472.3702	2	0.1396	none
					R.GCAIGDPAAYAR.V	611.3746	2	0.1739	none
					R.NVLIGVTSFGTGR.G	660.9694	2	0.2045	none
					R.VTSYISWINQR.L	683.9642	2	0.2098	none
McSP64	HM990186	3	15	258	R.IIGGELSPYER.N	617.3490	2	0.0433	none
					R.VTSYLSWISPQLIHK.-	591.3556	3	0.0782	none
					R.EVYVHPEYDVQFFTNDIAIGYLPR.R	962.5310	3	0.1641	none
McSP13	FJ205420	2	12	204	R.VLIGVTSFGSAR.G	603.9585	2	0.2257	none
					R.VTSFASWINAR.I	626.4506	2	0.2458	none
McSP59	HM990181	3	11	141	R.AAVQFNSPR.A	495.2560	2	-0.0116	none
					R.ISAYQNWLR.A	575.8033	2	-0.0010	none
					K.NDIGILITSSNVALNNLVR.T	676.0485	3	0.0019	none
McSP22	FJ205443	2	12	104	R.GTCSGDSGGPLVVSSNDK.Q	868.8732	2	-0.0366	none
					R.IATSDIVIHPNWTPSTAANDIAILR.I	897.1715	3	0.0691	none
McSP33	FJ205415	3	8	82	R.ASVELNVR.A	444.2795	2	0.0572	none
					R.ISAYQNWLR.S	575.8033	2	-0.0010	none
					R.SFLCGGSLITQR.T	669.8473	2	0.0038	none
2.2.2 Exopeptidases									
2.2.3.1 Aminopeptidases									
McAPN7	HM357836	34	37	5542	K.ATFDISITR.D	512.3091	2	0.0640	none
					R.FYQYFLTAK.V	590.8212	2	0.0315	none
					K.VEPEMGFETR.F	597.7897	2	0.0262	none
					R.VGAPLLEVMDR.Y	600.3526	2	0.0526	none
					K.VEPEMGFETR.F	605.7782	2	0.0082	Oxi. (M)
					R.VGAPLLEVMDR.Y	608.3339	2	0.0203	Oxi. (M)
					R.GNIPAGSGDYSLR.V	653.8387	2	0.0315	none
					R.GTTGNEWVIFNK.Q	683.3491	2	0.0112	none
					K.MSFTGILQSTMR.G	686.3845	2	0.0858	none

Appendix A. (continued)

Protein	Accession	#UP	SC (%)	MS	Peptide Sequence	m/z	z	Delta	Modification
					K.MSFTGILQSTMR.G	702.3518	2	0.0307	2 Oxi. (M)
					R.QAFPCYDEPGFK.A	729.8471	2	0.0510	none
					R.WMATTQFQPGHAR.Q	510.9234	3	0.0287	none
					R.WMATTQFQPGHAR.Q	516.257	3	0.0345	Oxi. (M)
					R.VDHTTGDMVVQQTR.F	534.9311	3	0.0245	Oxi. (M)
					R.DIEGFNPTISNMPIK.E	838.4430	2	0.0428	none
					R.DIEGFNPTISNMPIK.E	846.4364	2	0.0346	Oxi. (M)
					R.INVGTALQVGQQYIVK.M	577.6989	3	0.1021	none
					R.EALIIYDPENTNNFYK.Q	648.6611	3	0.0303	none
					R.DSNVAGSSSLWHIPLTWTR.G	709.6962	3	0.0124	none
					K.QAAIPDFSAGAMENWGLLTYR.E	771.0734	3	0.0882	none
					R.GGAPDFVNLKPSQILTAAQAVIPR.G	821.8188	3	0.0700	none
					R.ANILTYR.E	425.7541	2	0.0227	none
					R.ANIQFYNER.V	577.8008	2	0.0355	none
					R.EAFNILSFLK.F	591.3323	2	0.0008	none
					R.AQIVDDVFVLAR.A	673.4026	2	0.0505	none
					R.VFHWLQDNIER.T	486.2564	3	0.0216	none
					R.VNYDQTTWALITR.A	790.9103	2	0.0065	none
					K.ATPGVLNSMLSSITSR.L	817.4270	2	-0.0110	none
					K.ATPGVLNSMLSSITSR.L	825.4358	2	0.0116	Oxi. (M)
					R.LRDEIIDMSTAVVAR.L	563.6569	3	0.0562	none
					R.LRDEIIDMSTAVVAR.L	568.9772	3	0.0222	Oxi. (M)
					R.YGTAEDYDFFWSEYLK.S	1017.5858	2	0.2840	none
					K.FEDAYAPWLAAITGLNFAR.R	709.3799	3	0.0547	none
					K.SNVANEQVVMITAAGCTQHQPSLNR.F	909.1430	3	0.1001	none
McAPN3	HM357833	24	36	3209	R.YVISITR.D	426.2585	2	0.0112	none
					R.INTNPIDR.G	471.7722	2	0.0369	none
					R.GYYYVGTQLR.L	610.3312	2	0.0446	none
					K.AFPCFDEPQFK.S	693.3314	2	0.0360	none
					K.DMYAAFK.K	423.2107	2	0.0279	none
					R.AHEAFASYLR.G	582.8113	2	0.0357	none
					K.AVDEDFTYAR.D	593.7765	2	0.0082	none
					R.TIAFDVLDFLR.N	655.3803	2	0.0383	none
					R.VPDTGVMTVTQGR.Y	680.8623	2	0.0283	none
					R.VPDTGVMTVTQGR.Y	688.8568	2	0.0223	Oxi. (M)
					R.LYATTQFQPFHAR.K	527.2752	3	0.0095	none
					R.DINLGQSYSNMDIATTQQIGNR.V	813.7510	3	0.0815	none
					R.VTEPFTLDSYLELIHINFPAPILR.G	933.5400	3	0.0928	none
					R.DTVYPTDMNVDLDVDLVTASFSGIVR.M	953.5109	3	0.1417	Oxi. (M)
					K.AFIATIMAHELGHK.W	513.6155	3	0.0173	none
					R.AQIVNDVLHFVR.S	470.9447	3	0.0344	none
					R.NEHDYYVWNGALAQFDWILR.R	837.4285	3	0.0823	none
					R.DFGNINIETLFDSWVQNAGSPILTVNR.V	1007.5381	3	0.0885	none
					R.HVYCVGLR.E	502.2755	2	0.0296	none
					R.IVLNFLSTNR.A	588.9557	2	0.2307	none
					R.ATITSTYGGVDR.L	620.8048	2	-0.0146	none
					R.TNAFSFALQGNPENVR.I	588.9797	3	0.0583	none
					K.DQASLQHYLQQSMFNDR.I	694.3439	3	0.0667	none
McAPN6	HM357830	15	35	2665	R.FNINTGYSR.E	536.2601	2	-0.0089	none
					R.MTEHLLGSDVHR.Q	470.9121	3	0.0422	Oxi. (M)
					R.EKDFSTVKPIDLFNNLDLAGR.S	798.0886	3	0.0006	none
					R.SGIMSYER.A	479.7324	2	0.0276	Oxi. (M)
					R.TAGTEVPLDSR.N	573.3006	2	0.0143	none
					R.LHGKPELAELNR.R	459.5906	3	-0.0071	none
					R.AQIVNDVFQFAR.S	704.3679	2	-0.0093	none
					R.GNAEDFNFLWTR.F	735.3578	2	0.0274	none
					R.TFVDAIVQSNNLIR.R	795.4359	2	0.0000	none
					K.ILLLQTLGCTSDDASLR.T	625.9960	3	-0.0109	none
					R.YQLAPVLCNLGVESCLNAAYTQFDQLR.T	1048.5587	3	0.1327	none

Appendix A. (continued)

Protein	Accession	#UP	SC (%)	MS	Peptide Sequence	m/z	z	Delta	Modification
					R.VNYDNFTWDLITR.A	828.9099	2	0.0110	none
					R.TPLSYVSAR.L	497.2770	2	0.0103	none
					K.VFQYIQQNLAAVIHAFGNAR.T	754.0785	3	0.0225	none
					R.SILDTSYALVYSDAANTLESIR.L	801.4098	3	0.0062	none
					K.DQASLQHYLQQSMFNDR.I	699.6739	3	0.0619	Oxi. (M)
McAPN4	HM357829	10	20	725	R.LVSSVPSGDR.T	508.7826	2	0.0209	none
					R.QAFPSFDEPGFK.S	685.4484	2	0.2472	none
					K.DSLNSFVIHQNVR.E	510.3684	3	0.3039	none
					R.YVLTVSYIGNINETPLSR.G	680.4951	3	0.3899	none
					R.VALTFIR.N	410.2647	2	0.0135	none
					R.LQGCVTDAVTR.F	610.2871	2	-0.0429	none
					R.ALGCTITDTNR.Q	611.3019	2	0.0073	none
					R.AGGSDAWTFINNR.R	704.8558	2	0.0439	none
					R.QQYLTNILNDDIVK.A	838.9529	2	0.0132	none
					R.FQAHQTAPTVVANLVNPNLR.R	730.7435	3	0.0382	none
McAPN1	HM357837	5	11	263	K.TVTVSDDTGAEVK.L	661.3446	2	0.0338	none
					R.IYATTHFQPYNAR.Q	527.9428	3	0.0330	none
					R.YLQLTLTK.E	490.3118	2	0.0341	none
					R.AQVVDDVFALMR.S	682.3591	2	0.0070	none
					R.SSQTDIPQYSSVLSAITSAR.N	704.3693	3	0.0320	none
					R.YLQLTLTK.E	490.3118	2	0.0341	none
McAPN8	HM357839	2	5	202	K.YTEFEYEFPK.L	676.8367	2	0.0617	none
					K.LITSQLQPTFAR.R	687.9042	2	0.0272	none
2.2.3.2 Carboxypeptidases									
McCPA2	FJ210819	6	9	325	R.AIANYVNTIK.E	553.8368	2	0.0459	none
					R.DVGEFGFLLPR.Q	625.3642	2	0.0636	none
					K.GVAEVPIVYLFELR.D	535.6645	3	0.0742	none
					K.QDNPVIGMIEGGIHAR.E	569.6511	3	0.0746	none
					K.QDNPVIGMIEGGIHAR.E	574.9825	3	0.0737	Oxi. (M)
					K.KQDNPVIGMIEGGIHAR.E	612.3539	3	0.0880	none
McCPB4	FJ210817	7	21	205	R.YIDQVVR.E	446.7595	2	0.0230	none
					R.ETWEGIVVGAR.R	608.8392	2	0.0390	none
					K.YPDVATVVSPAK.S	623.8477	2	0.0204	none
					R.EWISPPTVTWAIHK.L	555.6214	3	-0.0299	none
					R.FDWILLPVVNPDGYK.F	888.4857	2	0.0274	none
					R.STNSNPLSQICPGVDGNR.N	639.2938	3	-0.0257	none
					R.QLPYDNYQELDVIDDYLDYIGEK.Y	941.1223	3	0.0433	none
McCPB1	FJ210818	4	15	163	R.SAVLSNIDR.T	487.7629	2	-0.0080	none
					K.AHDGADECPGVDGNR.N	523.8759	3	-0.0214	none
					R.YIVGNAASVLYYTTGTSR.D	646.0158	3	0.0519	none
					K.NSALANGIDWIIIPLANPDGYEYSIDEDR.M	1078.9075	3	0.1490	none
McCPA5	HM357840	2	11	134	R.YLSDYIDAISK.N	644.3477	2	0.0414	none
					R.RPTSTTTIGVDLSK.N	492.6161	3	0.0273	none
McCPA6	HM357841	1	5	61	R.EWVTLPATLYAIHK.L	548.0004	3	0.0866	none
2.2.3 Insect intestinal lipases									
McIIL4	HM357823	6	26	5273	K.GVVFGGLADK.T	481.6991	2	-0.1396	none
					R.LVDSDVVFHLFTR.I	774.4165	2	0.0041	none
					R.IVGVGLGGHIAGITAER.V	810.4687	2	0.0073	none
					R.IASFVNLLSSAFNYGPNNVR.I	1092.0549	2	-0.0218	none
					K.TGQSGIYSFQTNWQQPFAR.G	739.5040	3	0.4456	none
					R.FVGTACESYDEAIAQTCAGEK.G	769.6560	3	-0.0369	none
McIIL3	EU660855	7	31	663	K.YYLYTK.R	425.7274	2	0.0131	none
					R.VPLIGAEVAEFIK.W	693.4408	2	0.0705	none
					R.TDAEQTYSTASLR.V	721.8560	2	0.0289	none
					R.ACNSFTQVNLNLCR.G	566.2951	3	0.0814	none
					R.ITGLDPAGSGWGSNSQR.L	851.9460	2	0.0704	none
					R.AFEVFAASISNNNLVGR.A	905.0279	2	0.1196	none

Appendix A. (continued)

Protein	Accession	#UP	SC (%)	MS	Peptide Sequence	m/z	z	Delta	Modification
					K.EILLNRPNPSYNIIEVDWR.T	781.1232	3	0.1253	none
McIIL7	HM357826	2	6	177	R.APEFFAATVR.H	554.8941	2	0.2024	none
					R.HNSQVLVNGNANSVR.N	536.9462	3	0.0041	none
McIIL5	HM357824	2	9	169	R.APELFASSVR.T	538.8002	2	0.0196	none
					R.IANGDFYPNGGR.N	640.7890	2	-0.0311	none
McIIL6	HM357825	1	2	79	R.ATELFASTVR.H	547.7814	2	-0.0285	none
McIIL8	HM357827	1	8	53	K.TGSGLYHLVTNNR.Q	477.9100	3	-0.0185	none
McIIL9	HM357828	1	6	52	R.GLGQFLAFLNR.V	618.3496	2	0.0024	none
McIIL1	EU660853	1	7	42	R.SNPLVSQPIIQGNANLLAASNFNPAK.R	893.5373	3	0.1714	none
2.2.4 β-1, 3-Glucanase									
Mcβ1, 3-GLU	HM357842	1	4	55	R.FSIDDLELGR.V	582.7917	2	-0.0133	none
2.2.5 α-Amylase									
McAMY	HM357843	1	1	34	K.DLNQGTEYVR.Q	597.7956	2	0.0090	none
2.3 Other Enzymes									
2.3.1 Alkaline phosphates									
McALP1	HM357865	5	14	172	R.DAGIVTTTR.I	467.2725	2	0.0376	none
					R.AETESAFWTR.E	599.3002	2	0.0396	none
					R.ITHASPAGVFAK.V	400.2352	3	0.0333	none
					R.NVVMFLGDGMSVPTLAAAR.T	655.6881	3	0.0564	Oxi. (M)
					R.GNDILGPSR.D	464.7611	2	0.0302	none
2.3.2 dsRNASE									
McdsRNase	HM357845	11	22	838	R.NNSAFNWLR.W	561.3859	2	0.2159	none
					K.SDYVFATGQR.A	572.2703	2	-0.0096	none
					R.CWNNNLVIR.V	594.7997	2	-0.0021	none
					R.YLAANGNSGQIR.L	632.3407	2	0.0302	none
					R.WVDIYLVPQR.N	644.8477	2	-0.0168	none
					R.YITNHQFMAR.G	432.8693	3	-0.0221	Oxi. (M)
					R.NQIPVPLYFYK.V	691.5181	2	0.2774	none
					K.TVVTQYVGAALADR.Y	732.3937	2	-0.0051	none
					R.INGDLGQPQPVYLR.G	785.4116	2	-0.0225	none
					R.NIPHLPNFATNGLLA.-	796.4651	2	0.0637	none
					R.IGTAFISINNPHLTLAEAR.A	680.1719	3	0.3931	none
2.3.3 Astacin									
McAST	HM357847	2	7	62	R.AANEFNYVR.I	542.3289	2	0.1287	none
					K.YGTNIVNNLGLPYEYASNMHYGR.Y	888.1767	3	0.2800	Oxi. (M)
2.3.4 Pantetheinase									
McPAN	HM357848	2	8	76	R.SFSGVATGGTR.I	520.2720	2	0.0200	none
					R.FSSYIPDTTAIFEELTITAR.M	759.0764	3	0.0655	none
3. OTHER PROTEINS									
3.1 McMG176									
McMG176a	HM357849	5	32	613	R.GYGYYYLVR.I	577.3857	2	0.1964	none
					R.AGGVGFNQATIR.L	595.9233	2	0.2117	none
					K.NPIFNAVQYQDITYR.G	921.6464	2	0.3674	none
					R.KNPIFNAVQYQDITYR.G	657.4677	3	0.3756	none
					R.ITYLEAVEQGATQWGMPSLR.A	750.8609	3	0.4458	none
McMG176b	HM357850	4	32	512	R.GYGYYYLVR.I	577.3857	2	0.1964	none
					R.AGGVGFNQATIR.L	595.9233	2	0.2117	none
					K.NPIFNAVQYQDITYR.G	921.6464	2	0.3674	none
					R.ITYLEAVEQGATQWGIPTLR.A	749.5419	3	0.4297	none
McMG176c	HM357851	3	23	406	R.LVYNSEIR.A	497.3587	2	0.1738	none
					R.NLNQGSIGSGDR.L	609.4185	2	0.2429	none
					R.VGASQNATPSIIGGGIGSSFVTIR.I	763.8904	3	0.4368	none
McMG176e	HM357853	3	27	371	R.LVYNSEVR.A	490.3600	2	0.1921	none
					R.NLNQGSIGSGDR.L	609.4185	2	0.2429	none

Appendix A. (continued)

Protein	Accession	#UP	SC (%)	MS	Peptide Sequence	m/z	z	Delta	Modification
					R.VGASQNATPSIIGGGIGSSFVTIR.I	763.8904	3	0.4368	none
McMG176d	HM357852	2	10	115	R.GYGYYYLVR.I	577.3857	2	0.1964	none
					R.AGGVGFNHATIR.L	400.6056	3	0.1741	none
McMG176f	HM357854	3	27	60	R.NLNQGSIGSGDR.L	609.2934	2	-0.0075	none
					R.VGASQNATPSIIGGGIGSSFVTIR.I	763.8904	3	0.4368	none
					R.LVYNSEVR.A	490.3600	2	0.1921	none
3.2 Lipocalin									
McPOL	HM357855	20	23	4140	R.IGSWLLR.K	422.7658	2	0.0204	none
					R.GPLPAAATAR.I	462.7587	2	-0.0160	none
					R.HVGAWLLSR.T	519.8035	2	0.0154	none
					R.YLGTWLEVAR.Y	604.3201	2	-0.014	none
					R.YLGNWIEEAR.Y	625.8248	2	0.0259	none
					R.FADNIWYQIR.R	663.3233	2	-0.0244	none
					R.YPQSAQTGQCNR.A	705.3264	2	0.0229	none
					R.YPQTAQTGQCNR.A	712.3583	2	0.0711	none
					R.VPALSSAVSSYVTR.D	718.8573	2	-0.0670	none
					R.ATYAQGATAGVFTVENR.Q	878.4320	2	-0.0094	none
					R.TWPTAGNTALNNLVATR.Q	900.4789	2	0.0107	none
					R.DGDTYINVNFVEVTGEEVR.R	719.3442	3	0.0040	none
					R.LDDIMQFAQGTASPQTGAGTSAK.F	765.6978	3	-0.0132	none
					R.LDDIMQFAQGTASPQTGAGTSAK.F	771.0292	3	-0.0139	Oxi. (M)
					R.SLVELSEAANNAINTVISTTQGLR.E	834.4501	3	0.0152	none
					R.NLDAPLGTEATNNINAVITQTQGLK.E	866.1460	3	0.0658	none
					R.NLDVPLSIEATSNINSIITETQGLR.E	900.1796	3	0.0986	none
					R.RNLDAPLGTEATNNINAVITQTQGLK.E	918.1596	3	0.0056	none
					R.RNLDVPLSIEATSNINSIITETQGLR.E	952.1894	3	0.0267	none
					R.EEYYQDTAQTSNACFYYPVQADAEK.V	997.7554	3	-0.0108	none
3.3 REPAT									
McREPAT2	HM357858	2	9	86	R.LGEVLFYDR.L	556.3564	2	0.1274	none
					R.LVSLDSWPNAIK.V	671.9546	2	0.1654	none
3.4 Serpin									
McSERPIN1a	AY148483	2	24	42	K.GESDLFITDAIQK.A	718.8839	2	0.0338	none
					K.LLELPYEGDQSSLLIVLPNEIDGIGSLVEK.L	1085.2771	3	0.0714	none
3.5 C-type lectin									
McCLECT	HM357859	1	8	58	R.VEVNECGTVDPEYNLDKR.T	713.0202	3	0.0594	none
3.6 Lsti99 or Lsti201 like									
530_224_1	HM357857	1	3	779	K.SASNLDYIIVLAK.K	703.9379	2	0.0795	none
363_100_1	HM357856	3	15	322	K.VIIDSFR.Y	425.2917	2	0.0932	none
					R.GLCTGFTR.W	456.2361	2	0.0245	none
					R.INLFGTGIFSGYVTFLGVASAGSK.V	802.7626	3	0.0028	none
3.7 Soybean lectin									
GmLE1	K00821.1	3	7	73	K.VDENGTPKPSSLGR.A	486.2342	3	-0.051	none
					K.VLITYDASTSLLVASLVYPSQR.T	799.4078	3	-0.0983	none
					K.QPNMILQGDAIVTSSGK.L	887.927	2	-0.0536	Oxi. (M)
4. PROTEINS WITHOUT ORTHOLOGS									
357-94-1	HM357860	8	35	5959	K.ALAQFDR.A	410.7032	2	-0.0321	none
					K.DVEFFTR.T	457.2358	2	0.0229	none
					R.YSVPQTCALR.E	597.8070	2	0.0132	none
					K.ICTNFGDQVPR.G	653.8043	2	-0.0195	none
					K.ACVAQGGYLASSASSR.F	792.9052	2	0.0596	none
					R.ELMGNTFLQAQYLR.C	842.4599	2	0.0603	none
					R.ELMGNTFLQAQYLR.C	850.4559	2	0.0573	Oxi. (M)
					R.GCVHYSNQTDGFLAQK.A	608.9451	3	-0.0126	none
530-247-1	HM357861	4	13	576	K.ALAQFDR.A	410.7032	2	-0.0321	none
					K.DVDFFTR.T	450.2826	2	0.1322	none
					R.YSVPQTCAVR.E	590.8784	2	0.1716	none

Appendix A. (continued)

Protein	Accession	#UP	SC (%)	MS	Peptide Sequence	m/z	z	Delta	Modification
					R.ELLGNTFLQAQYLR.C	833.5831	2	0.2630	none
357-80-1	HM357862	1	4	50	R.ITSAPGVGLR.N	485.8455	2	0.1157	none

Abbreviations: #UP: Number of the unique peptides, SC: Sequence coverage, MS: Mowse score, Oxi: Oxidation.

	Amino acid composition by Peritrophin A Domain														Amino acid composition by Mucin Domain								Entire Protein			
	IIM1					IIM2		IIM3			IIM4				IIM1			IIM2	IIM3		IIM4		IIM1	IIM2	IIM3	IIM4
Residues	CBD1	CBD2	CBD3	CBD4	CBD5	CBD1	CBD2	CBD1	CBD2	CBD3	CBD1	CBD2	CBD3	CBD4	MD1	MD2	MD3	MD1	MD1	MD2	MD1	MD2				
Charged	**43.6**	**46.3**	**42.6**	**38.9**	**44.2**	**49.1**	**33.3**	**29.3**	**29.6**	**25.4**	**44.4**	**48.1**	**38.9**	**38.5**	**1.0**	**6.7**	**6.8**	**1.4**	**20.0**	**28.6**	**23.5**	**18.2**	**25.9**	**20.1**	**23.7**	**28.4**
Acidic	12.7	16.7	13.0	16.7	15.4	22.8	7.4	3.4	9.3	3.6	14.8	14.8	11.1	9.6	1.0	6.7	6.8	1.4	0.0	0.0	17.6	13.9	12.2	7.4	4.3	14.0
Basic	**5.4**	**5.6**	**7.4**	**1.8**	**5.8**	**8.8**	**3.7**	**5.2**	**0.0**	**1.8**	**5.6**	**9.3**	**3.7**	**5.8**	**0.0**	**0.0**	**0.0**	**0.0**	**5.0**	**0.0**	**5.9**	**0.8**	**2.8**	**4.5**	**2.4**	**3.7**
Polar	36.4	29.6	27.8	25.9	34.6	31.6	31.5	60.3	61.1	63.6	29.6	25.9	33.3	40.4	24.7	43.3	39.8	70.8	80.0	85.7	41.2	52.4	31.1	47.9	58.3	39.0
Hydpho	**25.4**	**31.5**	**31.5**	**33.3**	**30.8**	**26.3**	**37.0**	**22.4**	**20.4**	**23.6**	**33.7**	**25.9**	**27.8**	**28.8**	**48.4**	**30.0**	**37.7**	**9.7**	**10.0**	**14.3**	**17.6**	**17.1**	**32.4**	**22.3**	**26.5**	**24.6**
Ala	5.4	1.8	9.3	9.3	11.5	1.7	5.6	8.6	7.4	9.1	3.7	1.8	7.4	7.7	46.4	13.3	18.4	8.3	5.0	14.3	11.8	2.0	13.6	6.1	9.5	4.9
Arg	**1.8**	**1.8**	**1.8**	**0.0**	**1.9**	**3.5**	**3.7**	**3.4**	**0.0**	**1.8**	**0.0**	**5.6**	**1.8**	**1.9**	**0.0**	**0.0**	**0.0**	**0.0**	**5.0**	**0.0**	**5.9**	**0.0**	**1.2**	**2.6**	**1.4**	**1.5**
Asn	3.6	0.0	5.6	0.0	7.7	3.5	5.6	5.2	13.0	5.4	0.0	3.7	1.8	9.6	0.0	3.3	4.8	0.0	5.0	0.0	0.0	5.9	4.3	2.6	7.6	6.1
Asp	**1.8**	**7.4**	**5.6**	**7.4**	**5.8**	**14.0**	**5.6**	**3.4**	**5.6**	**3.6**	**7.4**	**5.6**	**5.6**	**3.8**	**0.0**	**5.0**	**1.9**	**0.0**	**0.0**	**0.0**	**0.0**	**7.1**	**5.7**	**4.5**	**3.3**	**6.3**
Cys	10.9	11.1	11.1	11.1	15.4	10.5	11.1	10.3	11.1	10.9	11.1	11.1	11.1	15.4	0.0	0.0	0.0	0.0	10.0	14.3	0.0	1.2	5.3	3.9	9.0	4.9
Gln	**1.8**	**1.8**	**3.7**	**3.7**	**0.0**	**5.3**	**5.6**	**3.4**	**0.0**	**3.6**	**1.8**	**5.6**	**3.7**	**1.9**	**0.0**	**0.0**	**0.0**	**0.0**	**5.0**	**0.0**	**5.9**	**0.0**	**1.4**	**2.6**	**2.8**	**1.7**
Glu	10.9	9.3	7.4	9.3	9.6	8.8	1.8	0.0	3.7	0.0	7.4	9.3	5.6	5.8	1.0	1.7	4.8	1.4	0.0	0.0	17.6	6.7	6.5	2.9	0.9	7.7
Gly	**5.4**	**1.8**	**3.7**	**5.6**	**5.8**	**1.7**	**3.7**	**3.4**	**3.7**	**3.6**	**1.8**	**3.7**	**7.4**	**7.7**	**1.0**	**0.0**	**1.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.8**	**5.1**	**2.3**	**2.8**	**4.8**
His	5.4	5.6	7.4	5.6	3.8	3.5	7.4	0.0	0.0	0.0	5.6	11.1	9.3	3.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.1	2.3	0.5	3.1
Ile	**3.6**	**3.7**	**5.6**	**1.8**	**1.9**	**3.5**	**3.7**	**0.0**	**5.6**	**0.0**	**1.8**	**5.6**	**3.7**	**5.8**	**0.0**	**3.3**	**1.0**	**0.7**	**0.0**	**0.0**	**5.9**	**4.4**	**2.8**	**2.6**	**3.8**	**4.3**
Leu	5.4	11.1	1.8	7.4	3.8	8.8	9.3	5.2	3.7	5.4	9.3	1.8	5.6	5.8	1.0	0.0	1.0	0.7	5.0	0.0	0.0	0.8	5.1	5.5	4.3	4.0
Lys	**3.6**	**3.7**	**5.6**	**1.8**	**3.8**	**5.3**	**0.0**	**1.7**	**0.0**	**0.0**	**5.6**	**3.7**	**1.8**	**3.8**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.8**	**1.6**	**1.9**	**0.9**	**2.1**
Met	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	1.3	0.9	0.5
Phe	**5.4**	**5.6**	**7.4**	**7.4**	**5.8**	**5.3**	**3.7**	**1.7**	**1.8**	**7.3**	**5.6**	**7.4**	**1.8**	**5.8**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**2.7**	**2.3**	**3.8**	**2.5**
Pro	9.1	9.3	9.3	11.1	3.8	5.3	9.3	5.2	5.6	3.6	9.3	9.3	7.4	3.8	24.7	20.0	14.6	18.1	5.0	0.0	17.6	15.1	12.9	12.0	4.3	10.4
Ser	**7.3**	**3.7**	**1.8**	**1.8**	**1.9**	**0.0**	**1.8**	**8.6**	**9.3**	**10.9**	**5.6**	**1.8**	**9.3**	**5.8**	**0.0**	**16.7**	**2.9**	**2.1**	**0.0**	**0.0**	**11.8**	**13.5**	**5.2**	**1.9**	**9.0**	**8.6**
Thr	3.6	5.6	1.8	5.6	5.8	8.8	3.7	22.4	18.5	23.6	3.7	1.8	3.7	3.8	24.7	23.3	32.0	68.7	55.0	57.4	23.5	29.4	12.4	34.9	22.3	14.9
Trp	**0.0**	**3.7**	**1.8**	**1.8**	**1.9**	**5.3**	**7.4**	**0.0**	**0.0**	**0.0**	**3.7**	**1.8**	**1.8**	**1.9**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**1.1**	**2.3**	**0.5**	**1.1**
Tyr	9.1	7.4	3.7	3.7	3.8	3.5	3.7	10.3	9.3	9.1	7.4	1.8	3.7	3.8	0.0	0.0	0.0	0.0	5.0	14.3	0.0	2.4	2.5	1.9	7.6	2.8
Val	**5.4**	**5.6**	**5.6**	**5.6**	**5.8**	**1.7**	**7.4**	**6.9**	**1.8**	**1.8**	**9.3**	**7.4**	**7.4**	**1.9**	**1.0**	**13.3**	**17.5**	**0.0**	**0.0**	**0.0**	**0.0**	**9.9**	**7.1**	**3.6**	**4.7**	**7.8**
Σ Res.	55	54	54	54	52	57	54	58	54	55	54	54	54	52	97	60	103	144	20	7	17	252	811	309	211	651

Values reported as percentage of the number of amino acids within the indicated region.

Appendix B. Amino acid composition of *M. configurata* insect intestinal mucins (McIIMs)

Protein	Type	Accession #	Unique peptides	Sequence Coverage (%)	Mowse Score	Peptide Sequence	m/z	z	Delta	Modification
LC-MS/MS from 1D gel										
McCDA1-32 kDa band	Chitin deacetylase	EU660852	1	3	80	K.IQIAHFANIPYESIK.G	872.6306	2	0.3111	none
McIIL1-32 kDa band	insect intestinal lipase	EU660853	1	8	81	K.TGSSGIYHVR.T	538.8773	2	0.1989	none
McIIL2-32 kDa band	insect intestinal lipase	EU660854	1	20	109	K.TGQSGIFYLETNSSPPFSR.G	1044.6861	2	0.3617	none
McIIL3-32 kDa band	insect intestinal lipase	EU660855	3	23	234	R.VPLIGAEVAEFIK.W	693.5211	2	0.2311	none
						R.TDAEQTYSTASLR.V	721.9651	2	0.2472	none
						R.AFEVFAASISNNNLVGR.A	905.1319	2	0.3275	none
McIIL2-30 kDa band	insect intestinal lipase	EU660854	1	20	145	K.TGQSGIFYLETNSSPPFSR.G	1044.6962	2	0.3819	none
McIIL3-30 kDa band	insect intestinal lipase	EU660855	1	7	80	R.TDAEQTYSTASLR.V	721.968	2	0.253	none
LC-MS/MS from 2D gel										
McCDA1-Loose fraction	Chitin deacetylase	EU660852	6	23	281	K.FILTNFER.H	520.2514	2	-0.0616	none
						R.CSSTNIPGGLQPR.D	693.8164	2	-0.0539	none
						K.EAEECTPETVCELPNCR.C	698.608	3	-0.0478	none
						R.APFGFFVHEAFLSAFPAVR.G	704.022	3	-0.0395	none
						K.IQIAHFANIPYESIK.G	872.4614	2	-0.0273	none
						R.DTPQFVTVTFDDGINVNNILTYR.N	881.4533	3	0.0093	none
McIIL1 isoform 1-spot 1	insect intestinal lipase	EU660853	2	17	63	K.TGSSGIYHVR.T	538.757	2	-0.0416	none
						K.AGNCSEEGTLK.L	583.2394	2	-0.0438	none
McIIL1 isoform 2-spot 2	insect intestinal lipase	EU660853	2	17	107	K.TGSSGIYHVR.T	538.7559	2	-0.0438	none
						K.AGNCSEEGTLK.L	583.2366	2	-0.0494	none
McIIL1 isoform 3-spot 3	insect intestinal lipase	EU660853	3	39	135	K.TGSSGIYHVR.T	538.7505	2	-0.0546	none
						K.AGNCSEEGTLK.L	583.2407	2	-0.0412	none
						R.SYFYMGESLATGGFTGTQCASLEEAK.A	941.0584	3	-0.0725	none
McIIL2 isoform 1-spot 4	insect intestinal lipase	EU660854	2	37	190	R.SFYYMAESLQSGGFTGR.R	958.8965	2	-0.0626	Oxidation (M)
						K.TGQSGIFYLETNSSPPFSR.G	1044.4681	2	-0.0743	none
McIIL2 isoform 1-spot 5	insect intestinal lipase	EU660854	1	17	74	R.SFYYMAESLQSGGFTGR.R	958.8975	2	-0.0606	Oxidation (M)

*Searches were performed using a fragment ion mass tolerance of 0.3 Da.

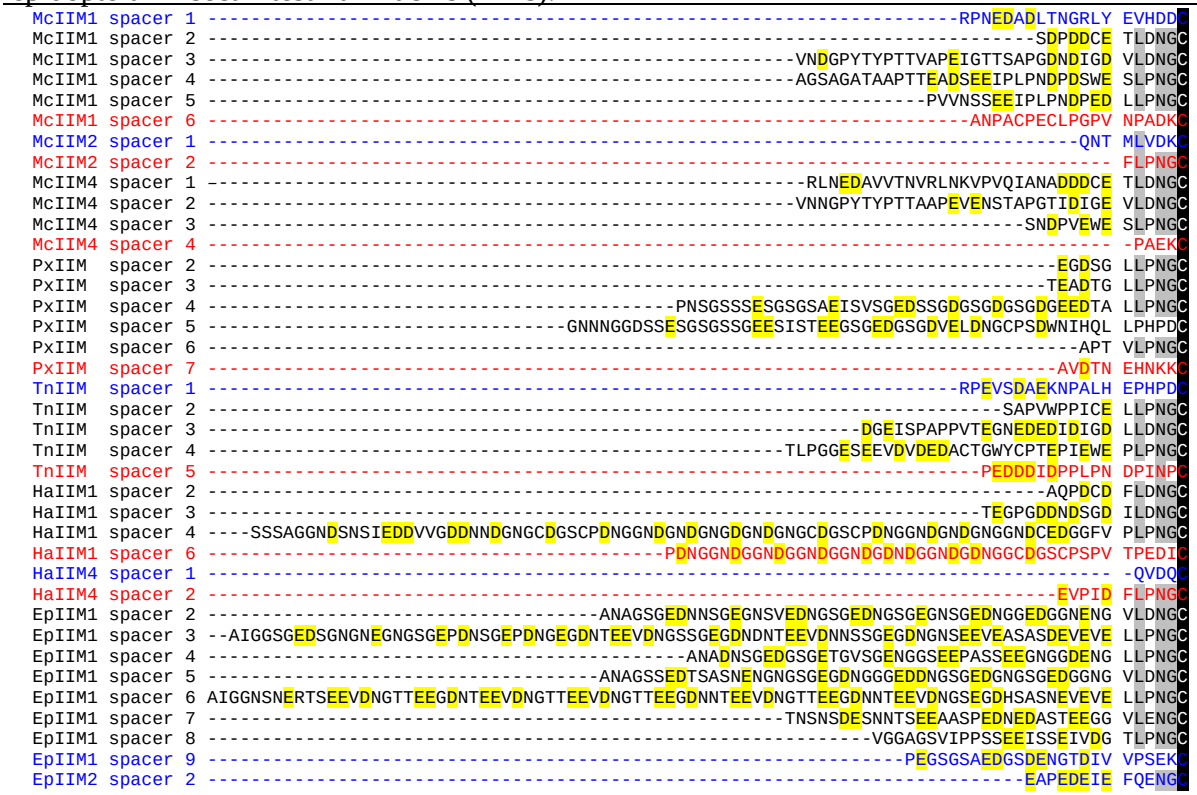
Appendix C. Data from *M. configurata* insect intestinal lipases 1-3 (McIIL1-3) and chitin deacetylase (McCDA1) identifications based on LC-MS/MS spectra

Protein	Type	Accession #	Unique peptides	Sequence Coverage (%)	Mowse Score	Peptide Sequence	m/z	z	Delta	Modification
McSP1	Trypsin	FJ205402	2	15	112	R.IVGGSVTTIDR.Y	559.4161	2	0.2037	none
						R.SASTFSFNNNVR.A	672.4461	2	0.251	none
McSP9*	Chymotrypsin	FJ205425	1	8	85	R.VGPCIGDSGGPLVFNR.S	823.0674	2	0.3113	none
McSP13	Chymotrypsin-elastase	FJ205420	2	28	70	R.VLIGVTSFGSAR.G	603.9598	2	0.2283	none
						R.GCQIGMPAAFAR.V	647.9415	2	0.2726	oxidation (M)
McSP23	Chymotrypsin	FJ205424	3	39	156	R.STCNGDSGGPLTVNR.S	767.9897	2	0.2807	none
						R.SGKPILIGITSFGSAR.G	802.6104	2	0.297	none
						R.GCQVGSPAAFAR.V	610.9073	2	0.2233	none
McSP27	Chymotrypsin	FJ205413	5	25	228	R.FTVVLGSVR.L	489.3856	2	0.1861	none
						R.RFTVVLGSVR.L	567.4423	2	0.1984	none
						R.ILIGVTSFGSDR.G	632.9628	2	0.2287	none
						R.GCQVGFPAAFAR.V	640.9336	2	0.2395	none
						R.STCQGDSGGPLVVTSNNR.R	925.0750	2	0.2923	none
McSP28	Trypsin	FJ205404	3	27	186	R.TADSSAGAITTSQGLR.H	768.5287	2	0.2842	none
						R.STCNGDSGGPLTVNR.S	767.9897	2	0.2807	none
						R.GCQVGSPAAFAR.V	610.9073	2	0.2233	none
McSP29*	Chymotrypsin	FJ205412	1	14	119	R.NSFPLIIQNSNICTSGAGGR.S	1053.7177	2	0.3885	none
McSP30	Chymotrypsin	FJ205436	2	11	75	R.IIGGTLASDTQFR.Y	689.9981	2	0.2564	none
						K.FTIHPNYR.T	524.3714	2	0.1985	none
McSP31	Trypsin	FJ205434	2	10	104	K.HPDYTEVPR.S	557.3723	2	0.2049	none
						K.DCVSDVYPDVFTR.I	787.0022	2	0.2972	none
McSP34*	Trypsin	FJ205445	2	16	128	R.SILTAAHCVIGDPAGR.W	819.5829	2	0.3158	none
						R.SASSIVYNNNVQR.A	726.4991	2	0.2672	none
McSP35	Chymotrypsin-elastase	FJ205405	2	22	95	R.NILIGVTSFGTGR.G	668.0012	2	0.2523	none
						R.VTSYISWINQR.L	683.9854	2	0.2522	none
McSP38*	Chymotrypsin	FJ205426	1	7	84	R.IHPDYNLASGENNIALININR.L	784.5513	3	0.4292	none
McSP41	Chymotrypsin	FJ205439	1	16	110	R.GGVGACIGDSGGPLILNR.S	857.0909	2	0.2997	none

* These proteins were identified when the mass tolerance in MASCOT searches was set to 0.6 Da whereas the others when mass tolerance was also set to 0.3 Da.

Appendix D. Data from serine proteases identified in 28 kDa band from *M. configurata* peritrophic matrix based on LC-MS/MS spectra

Appendix E. Alignment of spacer regions upstream of the chitin binding domains (CBDs) of lepidopteran insect intestinal mucins (IIMs).



GenBank accession numbers are: McIIM1 (AY057052), McIIM2 (FJ670567), McIIM4 (FJ670569); *Plutella xylostella* PxIIM (AF545582); *Trichoplusia ni* TnIIM (AF000606); *Helicoverpa armigera* HaIIM1 (EU325544), HaIIM4 (EU325565); *Epiphyas postvittana* EpIIM1 (EV809288, EV808201), EpIIM2 (EV809261, EV807804). Spacers nearest the amino terminus (blue), nearest the carboxy terminus (red) and internal to the protein (black) are shown. The regions are aligned relative to the first conserved cysteine in each CBD (black background) with a conserved motif (gray background) and negatively charged amino acids are indicated (yellow background).

Appendix F. Domain organization and sequences of lepidopteran proteins with peritrophin A domains (PPADs).

Mamestra configurata (Noctuidae)

McIIM1 (GenBank accession number, AY057052)

```
MIKTLLLVTA LALVQARPNE DADLTNGRLY EVHDDCPPAE VHFLLPHEYD CTKFYYCEYG LKYIEPRNCA SGTEFNAEIQ VCVHPSSSGC SLPGFSTLAP 100
GETAAPTAAP TAAPTAAPTA APTAAPTAAP TAAPTAAPTA APTAAPTAAP TAAPTAAPTA APTAAPTAAP TAAPTAAPTA APTAAPTAAP PVTSDPDDCE 200
TLDNGCPVDF TIHKLIPHEE YCHLFYYCDK GELLLRSCPQ PLYFDPATEV CVWSWETDCV NDGPYTYPTT VAPEIGTTSA PGDNDIGDVL DNGCPVDFSI 300
IHHLPHEECE KYYQCDAGKK IERNCAPGTV FNFAAQACDW PFNVPHCAGS AGATAAPTTE ADSEEIPLPN DPDSWESLPN GCPVDSSISH LVPHESDCDK 400
YYVCDNGRLV QLGCPAGTHF SPSQQFCTWP HEAGCEHWTG GGCTTPGNGG GSCGGSTAAP VDPTTPVAVV TSTSAPISDP STSAPNEPST PVAVVTTTTS 500
APISVPSTSA PNDPTTPVVN SSEEIPLPND PEDLLPNGCP ADFEVDLLLP HETDCDKFYY CVHGEIVEFP CAPGTHFSPA LQACTWPQEA GCEHWSEPST 600
VAPEITVTAV TSTLSVAPDT TAAVPNTPTV APETTTASVT NAPTVAPETT TAAVTNAPTV APETTTAVVT NAPTVAPETT TAVVTNAPTV APVPDPTTVG 700
TTANPACPEC LPGPVNPADK CKEECNVAPW AHAECDKYYT CVGDEFRVNA CAEGLHFNPS TLTCDFICNA GCVRNIPQIT RHVEGMLMFI PHDFNNRDVI 800
DLIEHELNAE L                                                                                                811
```

McIIM2 (FJ670567)

```
MKGIIILLFV VYAAVGRAQN TMLVDKCPEE QETDWTIELL LRHDDCNKFY KCTFGKPIEY TCPADLWFDL DTQQCDWRHN VDCTDRYVPG ETPTTEPTTA 100
APTTTEPTTL PTTTPTTTTT TTTTPAPTTT TSTTTTTTPA PTTTTSTTTT TTPAPTTTTT TTTTPAPTTT TTTTTPAPTT TPTTTTTTTT PAPTTTPTTT 200
TTTTTPAPTT TPTTTTTTTT PAPTTTTSTT AAPITFLPNG CPTNPHIHWL LPHESDCNAF YYCVWGQLVL RQCPATLHFN RVIQVCDWPW DAGCPVSLNK 300
HLDSRQMMR                                                                                                   309
```

McIIM3(FJ670568)

```
MAKQIIIIVF MLSICAWQAH AVVNCTATGA GRQADPADAT CKNYTLCVYV SSTNTYVSYN YVCPTTSLFS PTLRQCTTTY TCPATTTNTT TSVCTADGFI 100
ADPNSSNCSS YIECVNINGT YTETTYTCPA TTLYNPNTTL CDASYNCTST TAFTCTTAGR FANTADTTCQ TYFYCVLLSD GTFTQYNYTC PSTSSFNPAS 200
SLCTATYTCT N                                                                                                211
```

McIIM4(FJ670569)

```
MYKTLFFLTA LAIVQARLNE DAVVTNVRLN KVPVQIANAD DDCETLDNGC PVDFTIHKLV PHEEYCHLFY YCDKGELVLS SCPEPLYFDP KAQVCVWSWA 100
TDCVNNGPYT YPTTAAPEVE NSTAPGTIDI GEVLDNGCPS DIHIHHHLPH EECEKFYQCN FGQKVERDCA PGTVFHFEIQ VCDWPRNVPR CAGSAGATAR 200
PQTTPEASSE EIPTSNDPVE WESLPNGCPV DSSIHHLLPH ESVCDKYYAC DNGRLVEIGC ASGTHFSPAQ QVCTWPHEAG CEHWTGGGGC STPGDNGGSC 300
GGSTAVPISP TTPVSNISTT PVPDNITTTT TKTPETTTPV SDNSTTPVYE ITTTPVSDNS TTPVSDNSTT PVYEITTTPV SDNSTTPVSD NSTTPVYEIT 400
TTPVSDNSTT PVSDNSTTPV YEITTTPVSD NSTTPVSDNS TTPVYEITTT PVSDNSTTPV SDNSTTPVYE ITTTPVSDIE TTTTTKTPEV STTTTSAPDC 500
DTECTVTEDP ITEDSTPESN SSEEVPAPAD PELPGTTTNS PPAPSCPECP TVPLTPAEKC KQGCNVAPWA HAECDKYYSC IGNEFRLNIC SEGLHFNPST 600
LTCDFICNAG CDRNIPQVTR HEDGMLIFVP HSFNNKANML ELIEHELNEE F                                                     651
```

McPM1 (AY277403)

```
MKDAVILLLC AVALAQGLNQ SPDHRRPCNC DPSEAQQICQ ANYDNDDVLI AHENCDQFYK CANGKPVAYF CPNNLRYDPF SETCEWPDSV DCGNRPISDG 100
PDKGEDNDSD DVSDVDNDWT CNCNPGEAPS ICAAEGSNGI LVAHQNCNQF YKCAEGRPVT FDCSPTLLYN PYKEECDWAH NVECGDRVIP DLKEDDSSDD 200
DNNSTENDGT CNCNPEEAPA ICAAPGSESQ LIAHENCNKY YICNHGLPVA VSCVGDLLFN PYTRECDWPR NVDCGDRLVP ETECTGCNDN ASNDASCDGD 300
DDDSVPTPGV TNPAVTNPDP ESSDSGSSEI RPPGDDVVVP PRPPGTCNCN PGEAPSICAS EDSDGVLVAH ENCNQFYKCD HGKPVVLSCY GDLLYNPYTE 400
QCDWPENVDC GDRVIPDPDD SVITPGVTNP GVTNPGVTNP GVTNPADTTP GNNCDPSEAP AICAADDSEG VLVAHENCNQ FYMCSGGKPV ALKCPPNLLF 500
NPAKDKCDWP ENVDCGDRVV PDPESSDSGS SEIRPPGDDV VAPTRPPGTC NCNPGEAPSI CAAEDSDGVL VAHENCNQFY KCDHGKPVVL SCYGGLLYNP 600
YTEQCDWPEN VDCGDRVIPD PDDSVITPGV TNPGMTNPGV TNPGVTNPAD TTPGNNCDPS EAPAICAADD SEGVLVAHEN CNQFYMCSGG KPVALKCPPN 700
LLFNPAKDQC DWPENVDCGD RVIPDPESSD SGSSEIRPPG DDVVVPPRPP GTCNCNPGEA PSICASGDSD GVLVAHENCN QFYKCDHGKP VVLSCYGDLL 800
YNPYTEQCDW PENVDCGDRV IPDPDDSVIT PGVTNPGVTN PGVTNPGVTN PADTTPGNNC DPSEAPAICA ADDSEGVLVA HENCNQFYMC SGGKPVALKC 900
PPNLLFNPAK DQCDWPENVD CGDRVIPDPE SSDSGSSEIR PPGDDVVVPP RPPGTCNCNP GEAPSICASE DSDGVLVAHE NCNQFYKCDH GKPVVLSCYG 1000
DLLYNPYTEQ CDWPENVDCG DRVIPDPDDS VITPGVTNPG VTNPGVTNPG VTNPADTTPG NNCDPSEAPA ICAADDSEGV LVAHENCNQF YMCSGGKPVA 1100
LKCPPNLLFN PAKDQCDWPE NVDCGDRVIP DPESSDSGSS EIRPPGDDVV VPPRPPGTCN CNPGEAPSIC ASGDSDGVLV AHENCNQFYK CDHGKPVVLS 1200
CYGDLLYNPY TEQCDWPENV DCGDRVIPDP DDSVITPGVT NPGVTNPGVT NPGVTNPADT TPGNNCDPSE APAICAADDS EGVLVAHENC NQFYMCSGSK 1300
PVALKCPPNL LFNPAKDQCD WPENVDCGDR VIPDPESSDS GSSEIRPPGD DVVAPTRPPG TCNCNPGEAP SICAAEDSDG VLVAHENCNQ FYKCDHGKPV 1400
VLSCYGGLLY NPYTEQCDWP ENVDCGDRVI PDPDDSVITP GVTNPGVTNP GVTNPGVTNP ADTTPGNNCD PSEAPAICAA DDSEGVLVAH ENCNQFYKCS 1500
GGKPVALTCP PNLLFNPNKD QCDWPENVDC GDRVIPNPES SDSGSSEIRP PGDDVPPQPP VVDSNEDCSG ISDENGSPCN CDPDQAPSIC AVDNSEGVLI 1600
AHENCNQFYQ CVNGRPIPLK CPVNTLYNPV SQVCDWAFNV ECGDRIIPDP EENVSESNED DSKEEEPIVG PCNCNPEEAP AICAVDGSSG VQIAHENCNQ 1700
FYICDHGRPV AFTCNGFLLY NPYTERCDWP EHVQCGDRVI PEPGNESDEN DSNEDNISNP NDDPSQAPTI CAGNGSEGVL VAHENCDQYY ICSGGVPVSR 1800
PCNDGLLYNP YNQRCDWPSN VVCGDRIVPD DCACNPRNAP ALCAKPGSQG KLVAHENCNQ FYICSNSVPV SQTCPASLVY NPDREFCDWP QNVNCENRLL 1900
SYASLNKHWQ SRQTLRN                                                                                          1917
```

McPPAD1 (GU596430)

```
MIAKFLTTVL LLNVVLTAEI PQKNATETAK FRVFENVDPN DLSCDPAGHI FLLLPHFTDC SKFYMCAHGE EVEFQCPGGL IFDFQLQTCN WAWDTTCQLR 100
TPQDEDEGSG DEADSLIGIF TDELEHQPVD TVASVRPISP MQGRYNGIIN CARADAAARQ VPYKGDCQRY WKCVAGVPQV AFCSDGLFFN EHTQQCDFEA 200
NSKCVLSQED ELQSEFIKYE Q                                                                                     221
```

McPPAD2 (GU596431)

```
MSAKIVLVLC ALIAWVSSVP AQPGAGIGAW GGGGGGWGGA GGWGGAGGWG GAGAGAWGGA GAGVNPWIPP WVSSAPWYPQ WVPCSVVGSS CVDCNTKVVC 100
TKVGGLQRAC TDPTLPYCNA GECSATPSAV CGGPAVDAVA PRTAAVV                                                          147
```

McPPAD3 (GU596432)

```
MASVRARHLL ILLLVSCVKT DRVRKMESFN SAVFNSLREA SNQTQADTLN LPANASSIRE NITDTFSCEN RTYGYYADVD NDCQVFHVCL PSQTPSGRNI 100
TYRWSFICPN ETVFNQEVLT CTRVADSIDC KDSPDFYHVN LEIGKVSNKT EEVKDKKDMR KTQKRKQEKR NQNIVMGSEV SDFVTDDMQA HLEPVELIKR 200
QENGPVLIEA VIESVDPEIE DMEQEQKNEE EMDRIVMERN MRNDRQMRRG MFRFKGDV                                              258
```

McPPAD4 (GU596433)

```
MNTALWLLCC AAAVASSSGL SISLRQNPET LCRGQENFLR IASVESCFSY YQCYAGNSYL MQCPADTWFN EQEQICDWSS PPSHCVVPEP EPEPEPEPEP 100
QPEPEPEPEP EPEPEPEPEE EPEQEEEAEP TQELVEGSGE EWQEDAIRKY VKGFKGVKQE EIEAEDDETD EAKKLKQEED SDVEELEGEY LEENRKVQEE 200
IPEEAEEDLQ EKANKLMRQE EVEEEAADDE EAKKRQEEVE ADEEIEESKK KRQEEADDEE QADDDQEVKK RQEEIEIEDE VEEARERKRQE EVDGEESDDD 300
VETKKRQEEE SEDTNDEEEA KKRQEEDVEE SEDEVEGKKR QEEINDEEDV EKAKKRQEED FQESDDEVEA KKRQEVVEDE EENEEAKKRQ EEVNVEDESD 400
IEKTKKRQEE VLDDEESEDE VKRSIRQIEE EEENDIEEAK KRQEEVDGDE ESGDAIEPRK RQEEAEGDDE ANDDEANDDE ANDVEADDEE SKRSQEEDDN 500
EQAGDENDVE VEIDDETEVE VDDSLEARAK KFFKSVFRVL N                                                              541
```

Trichoplusia ni (Noctuidae)

TnIIM22 (AF000606)

```
MIKTLLFLTA LGLVAARPEV SDAEKNPALH EPHPDCPPAE QHWLLPHEYD CTKFYYCEYG LKFIAPRDCA PGTEFKFSAQ TCVHAALAGC TLPGPPAETT 100
QAPATTQAPT TTQAPTTTTQ APTTTTQAPT TTTQAPTTTQ APTTTQAPTT TQAPTTTTQA PTTTTQAPTT TTQAPTTTQA PTTTQAPTTT QAPTTITQAA 200
TTPAATTPAA TTPAATTPAA TTPAATTPGV PAPTSAPVWP PICELLPNGC PADFDIHLLI PHDKYCNLFY QCSNGYTFEQ RCPEGLYFNP YVQRCDSPAN 300
VECDGEISPA PPVTEGNEDE DIDIGDLLDN GCPANFEIDW LLPHGNRCDK YYQCVHGNLV ERRCGAGTHF SFELQQCDHI ELVGCTLPGG ESEEVDVDED 400
ACTGWYCPTE PIEWEPLPNG CPADFSIDHL LPHESDCGQY LQCVHGQTIA RPCPGNLHFS PATQSCESPV TACCQVFECD SDNQCTSTAA PTAAPTAAPT 500
AAPTAAPTAA PSTVVPPATP PATAAPVPPT TAIPTPAPTA APTAAPTTAA PESPTTVTVP PTAAPTAAPT TAVPEIPITV TSAPTAAPTA APTAAPTAAP 600
TTAVPEIPTT VTSPPTAAPT TAAPAPNTTV TVPPTAAPTT AAPAPNTTVT APPTAAPTTA APAPNTTVTV PPTAAPTAAP PTVAPAPNTT AAPVTTTSAP 700
ATTPEDDDID PPLPNDPINP CVEECNVLPW AHADCDKYWV CDGNNQVLVV CSEGLQFNPT TKTCDFACNV GCVRSNIQMS ESYEGVQVFI PWNKLDEDIR 800
QALNFEL                                                                                                    807
```

CBP1 (AY345124)

```
MKDQIILVLC ALALARGVDL DLKRQQCNCN PSEAQQICEA NYGSDNILIA HENCDKFYKC AHGKPEAITC PANLLYDPVL EVCNWAHEVD CGDRPISDGT 100
DQDCDGTTPG DVEDSPATCN CDPSEAPSIC AADGSNGALV AHENCNQFYI CSGGAPLPLV CPPGLLYNPY TTQCDWPQNV ECGDRVIPEL DDSPVTDNNN 200
DGNDDNNDGT CNCNPGEAPS ICAAPGSEGV LIAHENCNQY YICNFGRPIG FLCPGQLLYN PYLRVCDYPL NVDCGDRVVP EPENNCPSCN GGDGVDIDDV 300
PLQGEGNNNV GGGNCNPSEA PAICAANDSN DVLVAHENCN QYYICLGGKP IARPCPGNLL FNPISDQCDW PEKVNCGDRV IPDPDNDNSG NDNGGSDNGG 400
SDNGGSDNGG NDAIGGGNCD PSEAPAICAA DGSNDVLVAH ENCNKYYICL GGKPIARPCP GNLLFNPNSD QCDWPQNVDC GDRVIPDPDD DNSGNDNGGS 500
DNGGSDNGGG DNGGNDAIGG GNCDPSEAPA ICAADGSNDV LVAHENCNKY YICLGGKPIA RPCPGNLLFN PISDQCDWPE KVDCGERVIP DPDDDNNGND 600
NGGSDNGGSD NGSGDNGGSD SIGGGNCDPS EAPTICAADG SNDVLVAHEN CNKYYICLGG KPIARPCPGN LLFNPNSDQC DWPQNVDCGD RVIPDPDDDN 700
SGNDNGGSDN GGSDNGGNGG VCNCNPGEAP KICMAENSNG TLIAHENCDQ FYLCDHGKPV AMSCPDNLLY NPYKEECDWP VNVKCGDRNV PNSEDSEENK 800
EDIDNTFVVG DDQEDQVGAD CENDKLVNEE SCNCRPEDAP SICSVGGSDG VYIAHEYCNR FYQCSNGKPV AIRCPRGLLY NPVNITCDWP HNVECGDRII 900
PDSDEDDDVN ESEEVEDNEN DGTCNCNPEE APAICAAEGS SGVLVAHENC NQFYKCANGV PVAFTCSASL LYNPYRGDCD WPSNVECGNR PISVPDDNNV 1000
GTSTTTMPDD NQVINDDPSQ APSICAENGS SGVLVAHENC NQYYICSAGR PVPMPCSSGL LFNPVNRACD WPQNVVCGDR LIPEDDCACN PRNAPKLCSQ 1100
PNSDGRLVAH ENCNQFYFCA GGVPLVLSCP NSLLFNADQL ICDWPQNVNC NSRMFFAALN KHLESRQLLR K                               1171
```

Helicoverpa armigera (Noctuidae)

HaIIM86 (ABW98670)

```
MFKTLILLTA LASALARPSE EGSPNNGLLE YHPDCPPAEV HFLLPHEYDC TKFYYCEYGL KYIEPRDCPP GTEFSAEIQV CVHPASAGCT LPGYSTTSEA 100
PAPTTTAAPT TTAAPTTTTP EPTTTTTTPE PTTTTTTPEP TTTTTTTTPA PTTTTTTTTQ APTTTTTTTT QAPTTTTTTQ APTTTTTTTT QAPTTTTTTQ 200
APTTTTTTQA PTTTTTTQAP TTTTTTQAPT TTTAAPSTTA APSSTAVPSH CCPHCCTFHC CSIQHSPACC DFLDNGCPAD FDVHHLLPHE EYCNKFYYCV 300
FGEKVERDCA PGTWFRFEIP TCCWPYNVDC DKGVSEVTPT PQPPTIDRTG NDHDSGDILD NGCPADFDVH HLLPHETDCD KFYYCVFGEK VERICAPGTY 400
FNYEIQTCDW PHNVNCSSSG GGNDSNSIED DVVGDDNNDG NGCDGSCPDN GGNDGNDGNG DGNDGNGGCD GSCPDNGGND GNGDGNDCED GGFVPLPNGC 500
PADFDIHHLL PHEYDCNRFY YCVHGETVPA QCPPGTHFSV KEQICDWPHL AGCEDSGSGS CNGSCPDDGD NDGGNDGGND GGNGGCDGSC PDDGGNDGGN 600
DGDNDGGNDG GNGGCDGSCP DDGGNDGGND GDNDGGNDGD NDGGSDGGNG GCDGSCPDDG GNDGGNDGGN DGGNGDCDGD NDGGNDGGND GGDGGCDGSC 700
PDDGGNDGGN DGGNDGGNDG DNDGGNDGGD GGCDGSCPSP VTPEDICSNE CNVAPWAHDD CDKYWICKDS TPILITCSEG LHFNPSTYTC DFQCNAGCVR 800
ESIQTTADAS GVKIFLPWDK VNSEMKKLLK IQF                                                                        833
```

HaIIM2 (EU325543)

```
ATPAPTAAPT AAPTAGPTAA PTAAPTAAPT AAPTAAPTAA PTAAPTAAPT AAPTAAPTAA PTAAPTAAPT AAPTRPPIDE GTCNCNPGEA PSICAQDGSN 100
GTLIAHEDCN KFYICDHGKP VALSCPGNLL YNPYTEKCDW PENVECGDRV TPDPDATPAP TAAPTAAPTA GPTAGPTAAP TAAPTAGPTA APTAAPTRPP 200
IDEGTCNCNP GEAPSICAED DSDGILIAHE DCNKFYICDH GKPVVLSCPG DLLYNPYTEQ CDWPEKVECG DRLIPDPEQK PDPEDSGDDS SADIDDRPPP 300
GDDVTTRPPG TCNCNPEEAP SICAQDGSNG TLIAHEDCNK FYICDHGKPV ALSCPGNLLY NPYTEKCDWP ENVECGDRVT PDPDATPAPT AAPTAAPTAG 400
PTAGPTAAPT AAPTAGPTAA PTAAPTRPPI DEGTCNCNPG EAPSICAEDG SDGILIAHED CNKFYICDHG KPVVLSCPGD LLYNPYTEQC DWPENVECGD 500
RLIPDPDQKP DPEDSGDDSS ADIDDRPPPG DDVTTRPPGT CNCNPEEAPS ICAQDGSNGT LIAHEDCNKF YICDHGKPVA LSCPGNLLYN PYTEKCDWPE 600
NVECGDRVTP DPDATPAPTA APTAAPTAGP TAGPTAAPTA APTAGPTAAP TAAPTRPPID EGTCNCNPGE APSICAEDGS DGILIAHEDC NKFYICDHGK 700
PVVLSCPGNL LYNPYTEQCD WPENVECGDR VTPDPDATPA PTAAPTATPT ASPTAAPTAG PTAAPTAAPT AAPTTRPPID EGTCNCNPGE APSICAEDGS 800
DGILIAHEEC NKFYICDHGK PVVLSCPGNL FYNPYTEQCD WPVNVECGDR VTPDPDATPA PTAAPTSTTT VAPTAAPSTT TTVAPTAAPT TTTTVAPTTT 900
TTVAPTTTTT VAPTAAPTTT TTVAPTAAPT TTTTVASTAA PTAAPTAAPT AASTPDDSDC DDDNNGGDDT CNCNPDEAAS ICSVGNSDGI HVAHENCNWF 1000
YKCDNGRPVP FRCPSGLMYN PYTQICDWPW DVECGDRVIA DDDDSSEEDN DNDNDSGVVG PCNCNPEEAP AICAAEGSNG VHVAHQNCNQ YYMCDNGRPV 1100
AFTCNGFLLY NPYTQQCDWP HLVECGDRVI PEPGDEDDED CDDDDDNSNN VINDDPSQAP AICADSGSEG VLVAHENCDQ YYICDGGRPV ARPCQGGLLY 1200
NPLTQYCDWP GNVNCGDRII PDDCACNPRN APRLCSKPDS EGSLVAHENC NQFYICAHSV PVEHFCPVGL YYNIELELCD WAQNVNCENR NLPSLNKHWE 1300
SRQTLRK                                                                                                    1307
```

HaIIM3 (EU325564)

```
MFAKFLTITL TLTVAYAAET KLPDQKNTTD GNFKAFTDVD PNDLSCDPAG QVFLLLPHFT DCSKFFMCAH GEEVEFSCSG GLIFDFQLQT CNWPWATTCQ 100
LRTPKEEEEG SGDEADSLIG IFTDELEQQP VDMVASVRPI SPMLGRYNGI INCNRADAAA TQVPYKGDCQ RYWRCVAGVP QVAFCSDGLF FNSATQQCDF 200
EANSKCVLQQ EDELQSEFIK YEQ                                                                                  223
```

HaIIM4 (EU325565)

```
MKGIVILLLV ACAAVSRAQV DQCPSEQETD WEIEMLLRHE DCDKFYKCTF GKPIAYQCPS NLYFNLETWQ CDWRDNVDCT DRYDPGADPT TPKPTPEPPT 100
PEPTTPEPTT LEPTTPEPTT PEPTTPEPTT PEPTTTTTTP EPTTTTTTTT TTTPAPTTTT TTTTTPAPTT TTTTTTTTTP APTTAAPTEV PIDFLPNGCP 200
VNPHIHWLLP HETDCNAFYY CVWGQLVLRQ CPATLHFNRV IQVCDWPWDA GCPATLNKHL ESRVMFR                                   267
```

HaIIM5 (EU449965)

```
TPELTTTTPE LTTITPEQTT TTPELTTTTP EPTTTTPEAT TTTPELTTTT PELTTTPEPT TTTTAAPQLC PPGVFGTVPD PERCDAFYRC MGAVPPLHLT 100
CPTGFEYDPS VQNNCVLIAE GGCTLRKSAQ ASSFLNLNGT SETLDAQDID VEESTKLLAE IMCPVGSYGN VKDPACCESY YFCGGGTVMK MYCSQGFEFD 200
DVSKSCSPMS QNGCTAKLKG KLV                                                                                  223
```

Spodoptera frugiperda (Noctuidae)

SfPer-33k (contig assembled from AAS89976, contig 32 provided by Dr. W.R. Terra and Sf2M02530-5-1 from SpodoBase).

```
SPDPTTSPTG SPNPTPTPSP STPGSGTCNC RPEEAPSICA ADGSDGVLVA NENCNQFYIC DHGKPVALSC PGNLLYNPYI EQCDWPENVD CGDRVIPDTS 100
QTPPPGPSPD PTTSPTGSPN PTPTPSPSTP GSGTCNCRPE EAPSICAADG SDGVLVANEN CNQFYICAYG KPVALSCPGN LLYNPYIEQC DWPENVDCGD 200
RVVPDPSQTP PPGPGSSPNP TPSPTASTPS SGTCNCRPDE APSICTVDGS DGVLVAHENC NQFYKCDNGK PVALYCFGNL LYNPYTEQCD WPENVDCGDR 300
VIPDPGQTPA PGPGPSPSPT PSPSTPGSGT CNCRPDEAPS ICAVDGSDGV LVAHENCNQF YKCDNGKPVA LYCFGNLLYN PYTEQCDWPE NVDCGDRVIP 400
DPGQTPTPGP TPGPTPSPTP TPNPPGDNCD PSEAPTICAA DNSEGVLVAH ENCNQYYICS GSKPVAQTCP GNLLFNPSKD QCDWPENVDC GDRVIPDPGQ 500
TPIPSPSPTP SPSTPGSGTC NCRPDEAPSI CAVDGSDGVL VAHENCNQFY KCSDGKPVAL YCFGHLLYNP YTEQCDWPEN VDCGDRVIPD SSQTTSPTSC 600
RRPSPSTPGS GTCNCRPDEA PSICAVDGSD GVLVAHENCN QFYKCSDGKP VALYCFGHLL YNPYTEQCDW PENVDCGDRV IPDSSQSTSP TPAPSPTPAP 700
SPTTTPSPTP APSPTPAPSP TPAPSPDPES SESSDIDDLP KPDDNVSSPE DCSNSPDDNT CNCNPDQAPS ICAGANSNGI HIAHENCNQF YICNNGKPIP 800
FRCPSNLLYN PFIPGCDWAH NVDCGDRIIP DPDDTSEGPQ PTVPDDNNDN VGPGPCNHCN PEEAPAICAD ENSNGIHIAH QNCNQFFVCD HGRPVTFSCN 900
SLLLYNVYTK QCDWPSNVDC GDRVIPDRDI DSGNDSGENN NNNNEVYDDP SQAPTICAGS GSDGVLVAHE YCDQYYICDG GFPLSRPCHG SLLFNPQNQQ 1000
CDWPNNVNCG NRIVPDDCAC T                                                                                  1021
```

Sf27 (provided by Dr. W.R. Terra)

```
MKGIALLLLC IAAAQASPIC TIDMVGVMLP HPKCDHFYTC FFGSTAEQYC GDGLLFNPEL QVCDWPFNVD CGDRIVPPTK LIAEKTTKID NRSPAEICAE 100
EGSEGLILDH EYCDKYYKCN NGKPVTMPCP PNLLWSHVFC YWADRVDCGD RIRPEGFDQE PDDGVNDNLT PAEIFCSQLG TWKCIFAHEN CNHIYHAFFG 200
KPIVLSCPAI YSTT                                                                                         214
```

Spodoptera exigua (Noctuidae)

SeCBP66 (ABW98673)

```
MKGIALLLLC AAAAAQASSS CSPDSVGDLI PHRNCDQFYM CFFGSQTELH CADGLLFNPE AKVCDWPANV DCGDRIIPNR KLDVQKNLKK DGRSLSPAEI 100
CAAEDSEGLI FDHEYCDKYY KCNHGKPVTM PCPPNLLWYN PFCYWAAQVD CGDRIRPDGF NSGNQDAEKV PDNVSGGNSD PSQAPEICAA EDSDGILVAH 200
ENCNQFYKCF DGQPVALNCP ETLVFNPEKE YCDWSSNVEC GNRIIPDDIS NGNPNEDKGP ENVSEGNSDP SQAPAICAAE DSDGVLIAHE NCNQFYKCFD 300
GQPVALNCPE TLVFNPEKEY CDWSSNVECG NRIIPDDISN GNPNEDKGPE NVSEGNSDPS QAPAICAAED SDGVLIAHEN CNQFYKCLGG QPVAMNCPEN 400
LVYNPKREYC DWSWEVDCGN RIIPDDISSG KSNEDKSPEN VSGGNSDPSQ APEICAAEDS DGVLVAHENC NQFYKCLGGQ PVALNCPENL VYNPEREYCD 500
WSWEVDCGNR ILPDEIISEN SNEDKGPENV SGGNSDPSQA PEICAAEDSD GILVAHENCN QFYKCFDGQP VAMDCPENLF YNPGKEYCDW SSNVDCSNRI 600
KP                                                                                                      602
```

SeIIM (EU047712)

```
TTTTTQAPTT TTTTQAPTTT TTTQAPTTTT TTTQAPTTTT TTQAPTTTTQ ASTTTQQTSS VGPSSTSRPD CDFLDNGCPA DFDVHWLLPH EEYCNKFYYC 100
DKGQLVERFC APGTVFSPIV GVCVHPVDFD CGDKGIADQP DGPIDDNNKP CNGTDNDIGE VLENGCPANF DVHHLLPHET DCDKFYYCVH GQKVVSPCAP 200
GTHFNYEIQA CDWPYNVNCV PGGGDNDDDN DPDNEDDSDS ESVEINDPED NGFVPLPNGC PADFSIHHLL PHESDCSKFY YCVHGQKVVS SCGPGTHFNP 300
VLQVCDWPHN AGCEQSSNCP GGNNCPGDDD KPGCNGNCPE DDDKPGCNGN CPEDDDKPGC NGNCPGDDDD NGFVPLPNGC PSDFDIHHLP AHESDCSKFY 400
YCVHGQKVVS SCGPGTHFNP VLQVCDWPAN AGCEHWNPDG PVCDGNCPGD DDKPGCNGNC PEDDDKPGCN GNCPEDDDKP GCNGNCPEDD DKPGCNGNCP 500
EDDDKPGCNG NCPEDDDKPG CNGNCPEDDD KPGCNGNCPE DDDKPGCNGN CPEDDDKPGC NGNCPEDDDK PGCNGNCPED DDKPGCNGNC PEDDDKPGCN 600
GNCPGDDDPV CDGNCPGGDD PVTDNPCANQ CNVAPWASDD CDKYWMCVGD KKVQITCSEG LHFNPTTLTC DFHCNSGCER KKLELTEHYD GVRIFLPWNQ 700
LNSETKRLLS LEEL                                                                                         714
```

Epiphyas postvittana (Tortricidae)

EpIIM1 (EV809288)

```
YCVHGEKVQR ECGAGTHFNP VEQVCDWPQN AGCANAGSGE DNNSGEGNSV EDNGSGEDNG SGEGNSGEDN GGEDGGNENG VLDNGCPADF NIHQLLPHET 100
DCAKFYYCVH GEKVQRECGP GTHFNPAEQV CDWPHNAGCA IGGSGEDSGN GNEGNGSGEP DNSGEPDNGE GDNTEEVDNG SSGEGDNDNT EEVDNNSSGE 200
GDNGNSEEVE ASASDEVEVE LLPNGCPADF NIHQLLPHET DCAKFYYCVH GEKVERNCGT GTHFNPALQV CDHPENAGCA NADNSGEDGS GETGVSGENG 300
GSEEPASSEE GNGGDENGLL PNGCPADFNL HQLLPHETNC AKFYYCVHGE KVQRECGPGT HFNPVEQICD WPHNAGCANA GSSEDTSASN ENGNGSGEGD 400
NGGGEDDNGS GEDGNGSGED GGNGVLDNGC PADFNIHQLL PHETDCAKFY YCVHGEKVAR DCGEGTHFNP VEQICDWPHN AGCAIGGNSN ERTSEEVDNG 500
TTEEGDNTEE VDNGTTEEVD NGTTEEGDNN TEEVDNGTTE EGDNNTEEVD NGSEGDHSAS NEVEVELLPN GCPADFDIHQ LLPHETNCAK YYYCVHGEKQ 600
EREGNGELFF NPAIQVCDWP ANVDCTNSNS DESNNTSEEA ASPEDNEDAS TEEGGVLENG CPADFDIHQL LPHPYYCDRF FYCVNGAKVE RECGPGTHFN 700
PKLQVCDHPS NVNCVGGAGS VIPPSSEEIS SEIVDGTLPN GCPADFNIHQ LLPHETECDK FYYCVHGEKV ERNCGPGTHF NPTLQVCDHP ENAGCANAGS 800
NEDNETSPEA SDEVEPSPSS EESNPEGSGS VEIDPEDNGS AEIDPEDNGS AEIDPEDNSS AETTPEGSGS AEDGSDENGT DIVVPSEKCK SSCNIGSWAH 900
EHDCDKFWRC DGENAVQGTC SEGLHFNVKS QTCDFICNAG CVREDVQSTP EHDGLKLFLP WDKVDKTIRR LYLTGQK                      977
```

EpIIM2 (EV809261)

```
NIPGQEPETE APEPETEAPE PETEAPEPET EAPEPETEAP EPETEAPEPE TEAPEPETEA PEPETEAPEP ETEAPEPETE APEPETEAPE PETEAPEPET 100
EAPAPETEAP ATEAPEDEIE FQENGCPVDH SIHWLLPAED DCNGFYYCVW GEKVRRACAP TLHFNRVIQV CDWPADAGCV TSFNKHSEAR LTQGF      195
```

EpIIM3 (EV809248)

```
MLFALSILAL LSLASATQCP PHGSVRYMLP HETDCSKFYI CAHGVPLLQS CPEGTHFNNN FMVCQNPVFA GCSNELPASE LECSWPDFGH SRPHPDTCSK 100
FYVCEFGTPL VFDCPGGTEF NAELRECVWP HESNCNIGAQ TDAPEDTTDS IPLGDSGPEE SGPEDNGPQD GGSQDGRCVE GALSAHPTAV NTTSAITARR 200
SLWIARRVCS GAQN                                                                                         214
```

EpPTN1 (EV809224)

```
MKGAVLALFC AVALVHAEIN ENGCPINPFI DVLKAHPNCN KYYQCLQGEL FERKCASTLV FNVETQECDW PENVDCENRT PVAPESSEEE SQPEAGNGND 100
DPSLAPEICA EEGSDGILVA HENCNQFYEC NNGVPVAHRC ASTLLFNPAI NRCDWPDNVN CGDRNPDGGN GGEDGGNGDG GNGDGGTEDG GNGDGDNEDS 200
GNGDNEDGGN SDPSEAPEIC AAEGSDGVLV AHENCNQFYK CNGGVPVTFS CWGHLLFNPA NNQCDWPENV NCGDRNPDGG NGGGDGGNGD GGNGDGGNGD 300
GGNGDGGDGG NGGGNGNPSE APEICAAEGS DGVLVAHENC NQFYKCNGGV PVAMSCWGNL LFNPANDQCD WPENVNCEDR NPDGGNGGGD GGNGDGGNGD 400
GGNGDGGDGD SGSGGNGGGN GNPSEAPEIC AAEGSDGVLV AHENCNQFYV CNGGKPATLS CWGDLLFNPA LNQCDWPDNV NCERSPGNGG GDNGDNEDSG 500
NGGGNGDPSL APEICAADDS DSILVAHEDC NKFYVCSAGK PVTRNCYGDL HYNPSTESCD WPQNVDCGNR RTFTSFNKHL PLVRRM              586
```

EpPTN2 (EV806871)

```
MKGAVLALFC AVALVHAEIN ENGCPINPFI DVLKAHPNCN KYYQCLQGEL FERKCASTLV FNVETQECDW PENVDCENRT PVAPESSEEE SQPEAGNGND 100
DPSLAPEICA EEGSDGILVA HENCNQFYEC NNGVPVAHRC ASTLLFNPAI NRCDWPDNVN CGDRNPDGGN GGEDGGNGDG GNGDGGTEDG GNGDGDNEDS 200
GNGDNGDGGN SDPSEAPEIC AAEGSDGVLV AHENCNQFYK CNGGVPVTFS CWGHLLFNPA NNQCDWPENV NCGDRNPDGG NGGGDGGNGD GGNGDGGNGD 300
GGNGDGGNGD GGDGGNGGGN GNPSEAPEIC AAEGSDGVLV AHENCNQFYK CNGGVPVAMS CWGNLLFNPA NDQCDWPENV NCADRNPDGG NGGGDGGNGD 400
GGNGDGGTGD GGDGDSGSGG NGGGNGNPSE APEICAAEGS DGVLVAHENC NQFYKCNDGV PVAMNCWGNL LFNPANDQCD WPENVNCADR NPDGGNGGGD 500
GGSEDGGNGD GGSEDGGNGG GGSEDGGNGD GGNAGNNDPS LAPEICAAED SEGVLVAHEN CNQFYVCNGG KPVTLSCWGD LLFNPALNQC DWPDNVNCER 600
SPDNGGEDGG NGDGDNEDSG NGGGNGDPSL APEICAADDS DSILVAHEDC NKFYVCSAGK PVTRNCYGDL HYNPSTESCD WLQNVDCGNR RTFTSFNKHL 700
PLVRRM                                                                                                    706
```

Bombyx mori (Bombycidae)

BmIIM1 (BGIBMGA000185-PA)

```
MLKKFLVLTA LVIVQGRPQD ELTRRHEPSD LCPEKGHKLL PHEYDCTLFY YCEYGQKWST PRSCAPGTEF SAALQVCVHP AQADCNLPGR PTDSPPAEAP 100
TQAPTQAPTE TPTEPQTEPP TQPPTVAPTA PPTQAPAETP TEPQTEPPTQ PPTVAPTAPP TQAPAETPTE PQTELPTAAP TAAPTQAPVE PPTESPIPAP 200
PAPPIQPPVN PGETLPNGCP ADFHIHKLLP HEECNKFYYC VFGDKVERKC APGTYFNYKI QVCDHPWNVD CAGDGGNDGG DGGESGEDGG GGGGGGGGGG 300
GTLPNGCPAD FHIHQLLPHE YDCNRFYYCV HGEKVERQCS GDLHFNPVLQ YDCNRFYYCV HGEKVERRCS GDLHFNPVLQ VSGGGGGGGG GGGGGEGEGG 400
DGGGGGGGGG GEDGGGGGGG TLPNGCPADF DIHQLLPHEY DCNRFYYCVH GEKVERRCSG DLHFNPVLQV SSSMLYSIRI LLGGGGGGGG GGGGGGGGGG 500
GGGGGGGGGG GGGGGGGGGG GGGGGGGGGG GGGNNPSDKC KTECNILPWA HDTDCDKFWR CEGKEAVLVT CSEGLHFNPK KGSCDFICYA GCVRKTVQVT 600
TRPNGISIFV PLDNASQSEQ MTLITNSTK                                                                           629
```

BmIIM2 (BGIBMGA001480-PA)

```
MRGLILILAA VALARAENLC PPEQSENWEI ELLLPHPQCN KFYKCTFGQP VEMVCYGNLY FNLKTWQCDW PENVECGDRI DPADTTTASP TPTTLVPETT 100
TNRETEAPIT TPIPTTTETT TAEVTTTTTQ APTTPTLAPT TTTEVPTTTT QAPTTTTLAP TTTTEVPTTT TQAPTTTQEP DTEPPTTTAA PTVAPTPEPE 200
PEKPEIDFLE NGCPVNPHIH WLLPHEGNCN LFYYCVWGRK VLRHCPSTLH FNKVIQVCDW PWDAGCASSF DKNAVARRMI SEVEVSVNL            289
```

BmPPAD1 (BGIBMGA009641-PA)

```
MSSLGLLLRV VKTKGKPIEQ PKIFLFQVFL LLPHFRDCSK FYMCAHGEEV EFNCNGGLIF DFELQTCNWK WATNCTLRVV KEDIDTEGSG LGSGEETIGI 100
FGEEVENGPI DILTADSAGT VRPLSPSNLR FFNHELNCHR ADAAAKQVAY KGDCQRYWRC VGGVPQAMYC TDGLFFNELT QQCDFEANVK CGVIPDEELQ 200
GEFIVYKN                                                                                                  208
```

BmPPAD3 (BGIBMGA003270-PA)

```
MDKISNLIL ILMCGGLVMS YDEIDVEADT NIMLKILRDM ANQTKADTLN LPANATSIRE NITDTFSCEN RTYGYYADID NDCQVFHVCL PTQTPSGRAV 100
TYRYSFICPS ETVFNQEVLT CTRPRDAIEC EDSVLYYDLN EEIGKVPNKT EEQPPVNAQQ PAKSSSQQKE TETLPATNQR RTPNKRKQSL IIESLMKEVV 200
NDDLRQKLIG FAEQHTDDIP DSPEIPEELG LEDGAAVSDE EAEGRIALER RVGRKLDRGS FRFRADA                                  267
```

BmPPAD4 (BGIBMGA011851-PA)

```
MNKALWLLCC AAALVGASAF SINLRQNPET LCIGQENFLR IASIEDCFHY FQCYAGRSYL MRCPEETWFN EAEQVCDWSD LPERCAKVEP EPEPEPEPEP 100
EPEPEPEPEP EPEPEPEIEP EQEPEPESEP ELEPIDSEEN QVEEPAQPSD EVAENDEGSG EEWQEDAVWR FVKSSRVRQE EVEQEDDQDD EDEEGTDTVE 200
NDNDDNEEDD DENEEEEEAE KRSKRSLRQE ETEGDDENED DEDEDEDEEE VEKRSKRSLK QEETEDDDDD DDEEEDEDEE EVEKRSKRSL KQEETEDDDD 300
DDDEEGDEDE EEVEKRSKRS LKQEETEDDD DDDEEDEDE DEEEEEAEKR SKRSSRQEET EGDDDDEDDE DEDEEEVEKR SKRSLKQEET EVNDDDDYDD 400
EDGVEDEDEE EVEKRSKRSL RQEETEDNDN DEDDDDDEDD VEAEDEEEVE KRSKRSLRQE ETEDNDDDED EDDDDNEDDV NDEDEEEVEK RSKRSLRQEE 500
TENDDDDQEE IEDAEEELEA EEEELKEKSF IKRFFKFFA                                                                539
```

Plutella xylostella (Plutellidae)

PxIIM (AF545582)

```
ARATPEVTTP EVTTPEPEIT TTLAPETDAP VVVTTPPTTL EPETDAPEVV TPPATTLAPE TDAPEVVTPP ATTLAPETDA PEVVTATTEQ PEVVTAPSTT 100
LAPETDAPEV VTAPSTTLAP ETDAPEVVTA PSTTLAPETD APEVVTATTE VPEVVTAAST TEAPEVEKPT VVTAIPTTED PESETPEVVT AIPTTEAPEV 200
EKPTVVTAIP TTETTEDPTT LEPTTEGDSG LLPNGCPSDF HIHLLLPHET ECNLFYQCNF GEKVLKTCPK PLYFNNEIQV CDWPENVDCN GSNGGVTSPA 300
PTTEAETVEV VTAVPTTTES EAETVEVVTA IPTTTESEAE TVEVVTAIPT TNAPETEATT VAETETPEVV TAVPTTTESE AETVEVVTAI PTTTESEAET 400
VEVVTAIPTT NAPETEATTV AETETPEVVT AVPTTTESEA ETVEVVTAIP TTTESEAETV EVVTAIPTTN APETEATTVA ETKTPEVVTA VPTTTESEAE 500
TVEVVTAIPT TTESEAETVE VVTGTPTTAA PTTDAPVSTV TAVPITEAQT VEVVTATQTA EPETDATVTP VPTTTEAATT EADTGLLPNG CPADFHVHLL 600
LPHETECDLF YQCNFGEKVL KECPKPLLFN NELQVCDWEY NVECPNSGSS SESGSGSAEI SVSGEDSSGD GSGDGSGDGE EDTALLPNGC PADWSIHLLL 700
PHAECDKFYY CVHGNLVEHS CAPGTHFNPE IQVCDWPENV QCGNNNGGDS SESGSGSSGE ESISTEEGSG EDGSGDVELD NGCPSDWNIH QLLPHPDCDK 800
FYNCVHGNLV EQSCAPGTLF NPEIQVCDWP QNVQCGGTDK PEVVTAVPTT SEPEAETVEV VTSAPTTVTH EPTTEEAVVT VTATPEPIVT VTATEEPVAT 900
VTATQEPIVT VTATEEPVAT VTATEEPIVT VTATEEPVVT VTATQEPVVT VTATEEPIVT VTATQEPIVT VTATEEPVAT VTATQEPIVT 1000
VTATEEPVVT VTATQEPIVT VTATEEPVVT VTAPTVLPNG CPADSSIEQL LPHDSECGKF YQCVHGDLVE MACPIGLHFN PATERCDWPE SAGCAVDTNE 1100
HNKKCAEGCN VLPWAHETDC DKFYACDGQK ATLIVCAEGL HFNANTKTCD FICNANCARD NAQATAENDG VMIFLPWDKM DKDMLRKYGK QH         1192
```

Loxostege sticticalis (Pyralidae)

LsCBP (ACQ65651)

```
MKYTALILLS AVALAKSGDS GINPNGCPKD WSVEQLLPHP DCRKFYQCWD GKLVEHSCPE NLYFSVEANR CEWSEVVDCD ATYVPEKESD EADNNNIGAG 100
NCDPSKAPEI CAEEGSDSVL VAHEKCNKYY ICSHGKPVAL RCQVGLFFNP RTDRCDWPEN VDCGDRVIPD SGNGNNENGD GGNSNENNDN NQNVGAGNCD 200
PSKAPEICAA DDSDSVLVAH EKCNKYYICN HGKPVPMRCQ AGLFFNPKTD TCDWPENVDC GDRVIPDGGN GGNSNENNDN DDNGNGGNDN GNDNQNVGAG 300
NCDPSKAPEI CAADDSDSVL VAHEKCNKYY ICNHGIPVPR RCQAGLFFNP KTDTCDWPEN VDCGDRVIPD GGNGDNENGD GNNSNENNDN DDNGNGGNDN 400
GNDNQNVGAG NCDPSKAPEI CAADDSDSVL VAHEKCNKYY VCAHGKPVPM RCQAGLFFNP KTDTCDWPEN VDCGDRVISE GDNENDVGGN SNENNENDDN 500
GNGGNDNGGN DNGGNDNQNV GECNCDPGKA PEICAAEGSD NVLVAHEKCN KFYVCAHGRP FARRCQVGLL FNPKIERCDW PENVDCGDRL TSADENEGND 600
GNDVENDNVD NDNGGNDSQN TGSCNCDPSQ APKICAEDGS DNVLVAHEKC NKYYVCANGK PVSLRCPANL LYNPHKEICD WPENVECSEI VNPEIQDSED 700
GDSGDVNVGG GNNDPSLAPI ICADEKSDGV FVAHEICTKF YTCSNGKPVA LSCPASLFFN TSKDECDWPQ NVDCGNRRVP SGFDASLNKH LEARRTFMRW 800
V                                                                                                         801
```

Choristoneura fumiferana (Tortricidae)

Gasp (GQ144964)

```
MNSRCLLVLA AALGVAMAQE SFKCPDDFGF YPHHISCDKY WKCDNGVAEI KTCGNGLAFD ATDSKYLTEN CDYLHNVECG ERTQLEPPIS TPHCQRMYGI 100
FPDEQKCDVF WNCWNGEASR YQCSPGLAYD REARVCMWAD QVPECKNEEV ANGFACPAPG EVSNAGSFSR HAHPEDCRKY YICLEGVARE YGCPIGTVFK 200
IGDSDGTGNC EDPEDVPGCE DYYGDVDLKA LKKLGF                                                                   236
```

The colour scheme relates to the PAD clades in Figure 6.6 and Figure 6.7. Family names of the particular species ara given in paranthesis.

C....C, C....C, C....C, C....C, C....C, C....C: Six cysteine peritrophin A domain.

C....C : Eight cysteine carboxy terminal peritrophin A domain.

T/S....T/S : Mucin domain

C........C : Cysteine rich region

C : Additional cystein residues

M........ : Signal peptide

Appendix G. Amino acid sequences of individual peritrophin A domains used in the phylogenetic analysis.

1. McIIM1 CBD1
CPPAEVHFLLPHEYDCTKFYYCEYGLKYIEPRNCASGTEFNAEIQVCVHPSSSGC
2. McIIM1 CBD2
CPVDFTIHKLIPHEEYCHLFYYCDKGELLLRSCPQPLYFDPATEVCVWSWETDC
3. McIIM1 CBD3
CPVDFSIIHHLPHEECEKYYQCDAGKKIERNCAPGTVFNFAAQACDWPFNVPHC
4. McIIM1 CBD4
CPVDSSISHLVPHESDCDKYYVCDNGRLVQLGCPAGTHFSPSQQFCTWPHEAGC
5. McIIM1 CBD5
CPADFEVDLLLPHETDCDKFYYCVHGEIVEFPCAPGTHFSPALQACTWPQEAGC
6. McIIM1 CBD6
CKEECNVAPWAHAECDKYYTCVGDEFRVNACAEGLHFNPSTLTCDFICNAGC
7. McIIM2 CBD1
CPEEQETDWTIELLLRHDDCNKFYKCTFGKPIEYTCPADLWFDLDTQQCDWRHNVDC
8. McIIM2 CBD2
CPTNPHIHWLLPHESDCNAFYYCVWGQLVLRQCPATLHFNRVIQVCDWPWDAGC
9. McIIM3 CBD1
CTATGAGRQADPADATCKNYTLCVYVSSTNTYVSYNYVCPTTSLFSPTLRQCTTTYTC
10. McIIM3 CBD2
CTADGFIADPNSSNCSSYIECVNINGTYTETTYTCPATTLYNPNTTLCDASYNC
11. McIIM3 CBD3
CTTAGRFANTADTTCQTYFYCVLLSDGTFTQYNYTCPSTSSFNPASSLCTATYTC
12. McIIM4 CBD1
CPVDFTIHKLVPHEEYCHLFYYCDKGELVLSSCPEPLYFDPKAQVCVWSWATDC
13. McIIM4 CBD2
CPSDIHIHHHLPHEECEKFYQCNFGQKVERDCAPGTVFHFEIQVCDWPRNVPRC
14. McIIM4 CBD3
CPVDSSIHHLLPHESVCDKYYACDNGRLVEIGCASGTHFSPAQQVCTWPHEAGC
15. McIIM4 CBD4
CKQGCNVAPWAHAECDKYYSCIGNEFRLNICSEGLHFNPSTLTCDFICNAGC
16. McPM1 CBD1
CQANYDNDDVLIAHENCDQFYKCANGKPVAYFCPNNLRYDPFSETCEWPDSVDC
17. McPM1 CBD2
CAAEGSNGILVAHQNCNQFYKCAEGRPVTFDCSPTLLYNPYKEECDWAHNVEC
18. McPM1 CBD3
CAAPGSESQLIAHENCNKYYICNHGLPVAVSCVGDLLFNPYTRECDWPRNVDC
19. McPM1 CBD4
CASEDSDGVLVAHENCNQFYKCDHGKPVVLSCYGDLLYNPYTEQCDWPENVDC
20. McPM1 CBD5
CAADDSEGVLVAHENCNQFYMCSGGKPVALKCPPNLLFNPAKDKCDWPENVDC
21. McPM1 CBD6
CAAEDSDGVLVAHENCNQFYKCDHGKPVVLSCYGGLLYNPYTEQCDWPENVDC
22. McPM1 CBD7
CAADDSEGVLVAHENCNQFYMCSGGKPVALKCPPNLLFNPAKDQCDWPENVDC
23. McPM1 CBD8
CASGDSDGVLVAHENCNQFYKCDHGKPVVLSCYGDLLYNPYTEQCDWPENVDC
24. McPM1 CBD9
CAADDSEGVLVAHENCNQFYMCSGGKPVALKCPPNLLFNPAKDQCDWPENVDC
25. McPM1 CBD10
CASEDSDGVLVAHENCNQFYKCDHGKPVVLSCYGDLLYNPYTEQCDWPENVDC
26. McPM1 CBD11
CAADDSEGVLVAHENCNQFYMCSGGKPVALKCPPNLLFNPAKDQCDWPENVDC

27. McPM1 CBD12
CASGDSDGVLVAHENCNQFYKCDHGKPVVLSCYGDLLYNPYTEQCDWPENVDC
28. McPM1 CBD13
CAADDSEGVLVAHENCNQFYMCSGSKPVALKCPPNLLFNPAKDQCDWPENVDC
29. McPM1 CBD14
CAAEDSDGVLVAHENCNQFYKCDHGKPVVLSCYGGLLYNPYTEQCDWPENVDC
30. McPM1 CBD15
CAADDSEGVLVAHENCNQFYKCSGGKPVALTCPPNLLFNPNKDQCDWPENVDC
31. McPM1 CBD16
CAVDNSEGVLIAHENCNQFYQCVNGRPIPLKCPVNTLYNPVSQVCDWAFNVEC
32. McPM1 CBD17
CAVDGSSGVQIAHENCNQFYICDHGRPVAFTCNGFLLYNPYTERCDWPEHVQC
33. McPM1 CBD18
CAGNGSEGVLVAHENCDQYYICSGGVPVSRPCNDGLLYNPYNQRCDWPSNVVC
34. McPM1 CBD19
CAKPGSQGKLVAHENCNQFYICSNSVPVSQTCPASLVYNPDREFCDWPQNVNC
35. McPAD1 CBD1
CDPAGHIFLLLPHFTDCSKFYMCAHGEEVEFQCPGGLIFDFQLQTCNWAWDTTC
36. McPAD1 CBD2
CARADAAARQVPYKGDCQRYWKCVAGVPQVAFCSDGLFFNEHTQQCDFEANSKC
37. McPAD2 CBD1
CSVVGSSCVDCNTKVVCTKVGGLQRACTDPTLPYCNAGECSATPSAVC
38. McPAD3 CBD1
CENRTYGYYADVDNDCQVFHVCLPSQTPSGRNITYRWSFICPNETVFNQEVLTCTRVADSIDC
39. McPAD4 CBD1
CRGQENFLRIASVESCFSYYQCYAGNSYLMQCPADTWFNEQEQICDWSSPPSHC
40. TnIIM22 CBD1
CPPAEQHWLLPHEYDCTKFYYCEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGC
41. TnIIM22 CBD2
CPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVEC
42. TnIIM22 CBD3
CPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGC
43. TnIIM22 CBD4
CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGC
44. TnIIM22 CBD5
CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGC
45. TnCBP1 CBD1
CEANYGSDNILIAHENCDKFYKCAHGKPEAITCPANLLYDPVLEVCNWAHEVDC
46. TnCBP1 CBD2
CAADGSNGALVAHENCNQFYICSGGAPLPLVCPPGLLYNPYTTQCDWPQNVEC
47. TnCBP1 CBD3
CAAPGSEGVLIAHENCNQYYICNFGRPIGFLCPGQLLYNPYLRVCDYPLNVDC
48. TnCBP1 CBD4
CAANDSNDVLVAHENCNQYYICLGGKPIARPCPGNLLFNPISDQCDWPEKVNC
49. TnCBP1 CBD5
CAADGSNDVLVAHENCNKYYICLGGKPIARPCPGNLLFNPNSDQCDWPQNVDC
50. TnCBP1 CBD6
CAADGSNDVLVAHENCNKYYICLGGKPIARPCPGNLLFNPISDQCDWPEKVDC
51. TnCBP1 CBD7
CAADGSNDVLVAHENCNKYYICLGGKPIARPCPGNLLFNPNSDQCDWPQNVDC
52. TnCBP1 CBD8
CMAENSNGTLIAHENCDQFYLCDHGKPVAMSCPDNLLYNPYKEECDWPVNVKC
53. TnCBP1 CBD9
CSVGGSDGVYIAHEYCNRFYQCSNGKPVAIRCPRGLLYNPVNITCDWPHNVEC

54. TnCBP1 CBD10
CAAEGSSGVLVAHENCNQFYKCANGVPVAFTCSASLLYNPYRGDCDWPSNVEC
55. TnCBP1 CBD11
CAENGSSGVLVAHENCNQYYICSAGRPVPMPCSSGLLFNPVNRACDWPQNVVC
56. TnCBP1 CBD12
CSQPNSDGRLVAHENCNQFYFCAGGVPLVLSCPNSLLFNADQLICDWPQNVNC
57. PxIIM CBD2
CPSDFHIHLLLPHETECNLFYQCNFGEKVLKTCPKPLYFNNEIQVCDWPENVDC
58. PxIIM CBD3
CPADFHVHLLLPHETECDLFYQCNFGEKVLKECPKPLLFNNELQVCDWEYNVEC
59. PxIIM CBD4
CPADWSIHLLLPHAECDKFYYCVHGNLVEHSCAPGTHFNPEIQVCDWPENVQC
60. PxIIM CBD5
CPSDWNIHQLLPHPDCDKFYNCVHGNLVEQSCAPGTLFNPEIQVCDWPQNVQC
61. PxIIM CBD6
CPADSSIEQLLPHDSECGKFYQCVHGDLVEMACPIGLHFNPATERCDWPESAGC
62. PxIIM CBD7
CAEGCNVLPWAHETDCDKFYACDGQKATLIVCAEGLHFNANTKTCDFICNANC
63. HaIIM86 CBD1
CPPAEVHFLLPHEYDCTKFYYCEYGLKYIEPRDCPPGTEFSAEIQVCVHPASAGC
64. HaIIM86 CBD2
CPADFDVHHLLPHEEYCNKFYYCVFGEKVERDCAPGTWFRFEIPTCCWPYNVDC
65. HaIIM86 CBD3
CPGDFDVHHLLPHETDCDKFYYCVFGEKVERICAPGTYFNYEIQTCDWPHNVNC
66. HaIIM86 CBD4
CPADFDIHHLLPHEYDCNRFYYCVHGETVPAQCPPGTHFSVKEQICDWPHLAGC
67. HaIIM86 CBD5
CSNECNVAPWAHDDCDKYWICKDSTPILITCSEGLHFNPSTYTCDFQCNAGC
68. HaIIM2 CBD2
CAQDGSNGTLIAHEDCNKFYICDHGKPVALSCPGNLLYNPYTEKCDWPENVEC
69. HaIIM2 CBD3
CAEDDSDGILIAHEDCNKFYICDHGKPVVLSCPGDLLYNPYTEQCDWPEKVEC
70. HaIIM2 CBD4
CAQDGSNGTLIAHEDCNKFYICDHGKPVALSCPGNLLYNPYTEKCDWPENVEC
71. HaIIM2 CBD5
CAEDGSDGILIAHEDCNKFYICDHGKPVVLSCPGDLLYNPYTEQCDWPENVEC
72. HaIIM2 CBD6
CAQDGSNGTLIAHEDCNKFYICDHGKPVALSCPGNLLYNPYTEKCDWPENVEC
73. HaIIM2 CBD7
CAEDGSDGILIAHEDCNKFYICDHGKPVVLSCPGNLLYNPYTEQCDWPENVEC
74. HaIIM2 CBD8
CAEDGSDGILIAHEECNKFYICDHGKPVVLSCPGNLFYNPYTEQCDWPVNVEC
75. HaIIM2 CBD9
CSVGNSDGIHVAHENCNWFYKCDNGRPVPFRCPSGLMYNPYTQICDWPWDVEC
76. HaIIM2 CBD10
CAAEGSNGVHVAHQNCNQYYMCDNGRPVAFTCNGFLLYNPYTQQCDWPHLVEC
77. HaIIM2 CBD11
CADSGSEGVLVAHENCDQYYICDGGRPVARPCQGGLLYNPLTQYCDWPGNVNC
78. HaIIM2 CBD12
CSKPDSEGSLVAHENCNQFYICAHSVPVEHFCPVGLYYNIELELCDWAQNVNC
79. HaIIM3 CBD1
CDPAGQVFLLLPHFTDCSKFFMCAHGEEVEFSCSGGLIFDFQLQTCNWPWATTC
80. HaIIM3 CBD2
CNRADAAATQVPYKGDCQRYWRCVAGVPQVAFCSDGLFFNSATQQCDFEANSKC

81. HaIIM4 CBD1
CPSEQETDWEIEMLLRHEDCDKFYKCTFGKPIAYQCPSNLYFNLETWQCDWRDNVDC
82. HaIIM4 CBD2
CPVNPHIHWLLPHETDCNAFYYCVWGQLVLRQCPATLHFNRVIQVCDWPWDAGC
83. HaIIM5 CBD2
CPPGVFGTVPDPERCDAFYRCMGAVPPLHLTCPTGFEYDPSVQNNCVLIAEGGC
84. HaIIM5 CBD3
CPVGSYGNVKDPAGCESYYFCGGGTVMKMYCSQGFEFDDVSKSCSPMSQNGC
85. EpIIM1 CBD3
CPADFNIHQLLPHETDCAKFYYCVHGEKVQRECGPGTHFNPAEQVCDWPHNAGC
86. EpIIM1 CBD4
CPADFNIHQLLPHETDCAKFYYCVHGEKVERNCGTGTHFNPALQVCDHPENAGC
87. EpIIM1 CBD5
CPADFNLHQLLPHETNCAKFYYCVHGEKVQRECGPGTHFNPVEQICDWPHNAGC
88. EpIIM1 CBD6
CPADFNIHQLLPHETDCAKFYYCVHGEKVARDCGEGTHFNPVEQICDWPHNAGC
89. EpIIM1 CBD7
CPADFDIHQLLPHETNCAKYYYCVHGEKQERECNGELFFNPAIQVCDWPANVDC
90. EpIIM1 CBD8
CPADFDIHQLLPHPYYCDRFFYCVNGAKVERECGPGTHFNPKLQVCDHPSNVNC
91. EpIIM1 CBD9
CPADFNIHQLLPHETECDKFYYCVHGEKVERNCGPGTHFNPTLQVCDHPENAGC
92. EpIIM1 CBD10
CKSSCNIGSWAHEHDCDKFWRCDGENAVQGTCSEGLHFNVKSQTCDFICNAGC
93. EpIIM2 CBD2
CPVDHSIHWLLPAEDDCNGFYYCVWGEKVRRACAPTLHFNRVIQVCDWPADAGC
94. EpIIM3 CBD1
CPPHGSVRYMLPHETDCSKFYICAHGVPLLQSCPEGTHFNNNFMVCQNPVFAGC
95. EpIIM3 CBD2
CSWPDFGHSRPHPDTCSKFYVCEFGTPLVFDCPGGTEFNAELRECVWPHESNC
96. EpPTN1 CBD1
CPINPFIDVLKAHPNCNKYYQCLQGELFERKCASTLVFNVETQECDWPENVDC
97. EpPTN1 CBD2
CAEEGSDGILVAHENCNQFYECNNGVPVAHRCASTLLFNPAINRCDWPDNVNC
98. EpPTN1 CBD3
CAAEGSDGVLVAHENCNQFYKCNGGVPVTFSCWGHLLFNPANNQCDWPENVNC
99. EpPTN1 CBD4
CAAEGSDGVLVAHENCNQFYKCNGGVPVAMSCWGNLLFNPANDQCDWPENVNC
100. EpPTN1 CBD5
CAAEGSDGVLVAHENCNQFYVCNGGKPATLSCWGDLLFNPALNQCDWPDNVNC
101. EpPTN1 CBD6
CAADDSDSILVAHEDCNKFYVCSAGKPVTRNCYGDLHYNPSTESCDWPQNVDC
102. EpPTN2 CBD1
CPINPFIDVLKAHPNCNKYYQCLQGELFERKCASTLVFNVETQECDWPENVDC
103. EpPTN2 CBD2
CAEEGSDGILVAHENCNQFYECNNGVPVAHRCASTLLFNPAINRCDWPDNVNC
104. EpPTN2 CBD3
CAAEGSDGVLVAHENCNQFYKCNGGVPVTFSCWGHLLFNPANNQCDWPENVNC
105. EpPTN2 CBD4
CAAEGSDGVLVAHENCNQFYKCNGGVPVAMSCWGNLLFNPANDQCDWPENVNC
106. EpPTN2 CBD5
CAAEGSDGVLVAHENCNQFYKCNDGVPVAMNCWGNLLFNPANDQCDWPENVNC
107. EpPTN2 CBD6
CAAEDSEGVLVAHENCNQFYVCNGGKPVTLSCWGDLLFNPALNQCDWPDNVNC

108. EpPTN2 CBD7
CAADDSDSILVAHEDCNKFYVCSAGKPVTRNCYGDLHYNPSTESCDWLQNVDC
109. SfPer-33k CBD2
CAADGSDGVLVANENCNQFYICDHGKPVALSCPGNLLYNPYIEQCDWPENVDC
110. SfPer-33k CBD3
CAADGSDGVLVANENCNQFYICAYGKPVALSCPGNLLYNPYIEQCDWPENVDC
111. SfPer-33k CBD4
CTVDGSDGVLVAHENCNQFYKCDNGKPVALYCFGNLLYNPYTEQCDWPENVDC
112. SfPer-33k CBD5
CAVDGSDGVLVAHENCNQFYKCDNGKPVALYCFGNLLYNPYTEQCDWPENVDC
113. SfPer-33k CBD6
CAADNSEGVLVAHENCNQYYICSGSKPVAQTCPGNLLFNPSKDQCDWPENVDC
114. SfPer-33k CBD7
CAVDGSDGVLVAHENCNQFYKCSDGKPVALYCFGHLLYNPYTEQCDWPENVDC
115. SfPer-33k CBD8
CAVDGSDGVLVAHENCNQFYKCSDGKPVALYCFGHLLYNPYTEQCDWPENVDC
116. SfPer-33k CBD9
CAGANSNGIHIAHENCNQFYICNNGKPIPFRCPSNLLYNPFIPGCDWAHNVDC
117. SfPer-33k CBD10
CADENSNGIHIAHQNCNQFFVCDHGRPVTFSCNSLLLYNVYTKQCDWPSNVDC
118. SfPer-33k CBD11
CAGSGSDGVLVAHEYCDQYYICDGGFPLSRPCHGSLLFNPQNQQCDWPNNVNC
119. Sf27 CBD2
CTIDMVGVMLPHPKCDHFYTCFFGSTAEQYCGDGLLFNPELQVCDWPFNVDC
120. Sf27 CBD3
CAEEGSEGLILDHEYCDKYYKCNNGKPVTMPCPPNLLWSHVFCYWADRVDC
121. BmIIM1 CBD1
CPEKGHKLLPHEYDCTLFYYCEYGQKWSTPRSCAPGTEFSAALQVCVHPAQADC
122. BmIIM1 CBD2
CPADFHIHKLLPHEECNKFYYCVFGDKVERKCAPGTYFNYKIQVCDHPWNVDC
123. BmIIM1 CBD3
CPADFHIHQLLPHEYDCNRFYYCVHGEKVERQCSGDLHFNPVLQYDCNRFYYC
124. BmIIM1 CBD4
CKTECNILPWAHDTDCDKFWRCEGKEAVLVTCSEGLHFNPKKGSCDFICYAGC
125. BmIIM2 CBD1
CPPEQSENWEIELLLPHPQCNKFYKCTFGQPVEMVCYGNLYFNLKTWQCDWPENVEC
126. BmIIM2 CBD2
CPVNPHIHWLLPHEGNCNLFYYCVWGRKVLRHCPSTLHFNKVIQVCDWPWDAGC
127. BmPAD1 CBD1
CDPAGHVFLLLPHFRDCSKFYMCAHGEEVEFNCNGGLIFDFELQTCNWKWATNC
128. BmPAD1 CBD2
CHRADAAAKQVAYKGDCQRYWRCVGGVPQAMYCTDGLFFNELTQQCDFEANVKC
129. BmPAD3 CBD1
CNRTYGYYADIDNDCQVFHVCLPTQTPSGRAVTYRYSFICPSETVFNQEVLTCTRPRDAIEC
130. BmPAD4 CBD1
CIGQENFLRIASIEDCFHYFQCYAGRSYLMRCPEETWFNEAEQVCDWSDLPERC
131. SeCBP66 CBD1
CSPDSVGDLIPHRNCDQFYMCFFGSQTELHCADGLLFNPEAKVCDWPANVDC
132. SeCBP66 CBD2
CAAEDSEGLIFDHEYCDKYYKCNHGKPVTMPCPPNLLWYNPFCYWAAQVDC
133. SeCBP66 CBD3
CAAEDSDGILVAHENCNQFYKCFDGQPVALNCPETLVFNPEKEYCDWSSNVEC
134. SeCBP66 CBD4
CAAEDSDGVLIAHENCNQFYKCFDGQPVALNCPETLVFNPEKEYCDWSSNVEC

135. SeCBP66 CBD5
CAAEDSDGVLIAHENCNQFYKCLGGQPVAMNCPENLVYNPKREYCDWSWEVDC
136. SeCBP66 CBD6
CAAEDSDGVLVAHENCNQFYKCLGGQPVALNCPENLVYNPEREYCDWSWEVDC
137. SeCBP66 CBD7
CAAEDSDGILVAHENCNQFYKCFDGQPVAMDCPENLFYNPGKEYCDWSSNVDC
138. SeIIM CBD2
CPADFDVHWLLPHEEYCNKFYYCDKGQLVERFCAPGTVFSPIVGVCVHPVDFDC
139. SeIIM CBD3
CPANFDVHHLLPHETDCDKFYYCVHGQKVVSPCAPGTHFNYEIQACDWPYNVNC
140. SeIIM CBD4
CPADFSIHHLLPHESDCSKFYYCVHGQKVVSSCGPGTHFNPVLQVCDWPHNAGC
141. SeIIM CBD5
CPSDFDIHHLPAHESDCSKFYYCVHGQKVVSSCGPGTHFNPVLQVCDWPANAGC
142. SeIIM CBD6
CNVAPWASDDCDKYWMCVGDKKVQITCSEGLHFNPTTLTCDFHCNSGC
143. LsCBP CBD1
CPKDWSVEQLLPHPDCRKFYQCWDGKLVEHSCPENLYFSVEANRCEWSEVVDC
144. LsCBP CBD2
CAEEGSDSVLVAHEKCNKYYICSHGKPVALRCQVGLFFNPRTDRCDWPENVDC
145. LsCBP CBD3
CAADDSDSVLVAHEKCNKYYICNHGKPVPMRCQAGLFFNPKTDTCDWPENVDC
146. LsCBP CBD4
CAADDSDSVLVAHEKCNKYYICNHGIPVPRRCQAGLFFNPKTDTCDWPENVDC
147. LsCBP CBD5
CAADDSDSVLVAHEKCNKYYVCAHGKPVPMRCQAGLFFNPKTDTCDWPENVDC
148. LsCBP CBD6
CAAEGSDNVLVAHEKCNKFYVCAHGRPFARRCQVGLLFNPKIERCDWPENVDC
149. LsCBP CBD7
CAEDGSDNVLVAHEKCNKYYVCANGKPVSLRCPANLLYNPHKEICDWPENVEC
150. LsCBP CBD8
CADEKSDGVFVAHEICTKFYTCSNGKPVALSCPASLFFNTSKDECDWPQNVDC

	Amino acid composition by Peritrophin A Domain																								Entire Protein			
	PM1																				PPAD1		PPAD2	PPAD3	PM1	PPAD1	PPAD2	PPAD3
Residues	CBD1	CBD2	CBD3	CBD4	CBD5	CBD6	CBD7	CBD8	CBD9	CBD10	CBD11	CBD12	CBD13	CBD14	CBD15	CBD16	CBD17	CBD18	CBD19		CBD1	CBD2	CBD1	CBD1				
Charged	**44.4**	**39.6**	**37.8**	**45.3**	**37.8**	**43.4**	**35.9**	**43.4**	**35.9**	**45.3**	**35.9**	**43.4**	**35.9**	**43.4**	**35.9**	**30.2**	**37.7**	**34.0**	**30.2**		**33.3**	**38.9**	**31.2**	**36.5**	**33.2**	**32.1**	**13.6**	37.2
Acidic	18.5	13.2	11.3	18.9	15.1	17.0	15.1	17.0	15.1	18.9	15.1	17.0	15.1	17.0	15.1	9.4	11.3	9.4	7.6		13.0	11.1	6.3	12.7	16.3	15.4	2.7	17.1
Basic	**5.6**	**5.7**	**5.7**	**3.8**	**7.6**	**3.8**	**5.7**	**3.8**	**5.7**	**3.8**	**5.7**	**3.8**	**5.7**	**3.8**	**5.7**	**3.8**	**3.8**	**3.8**	**5.7**		**1.9**	**11.1**	**6.3**	**6.4**	**4.7**	**6.8**	**3.4**	**14.0**
Polar	38.9	37.8	35.9	37.8	28.3	35.9	30.2	37.8	30.2	37.8	30.2	37.7	32.1	35.9	34.0	39.6	37.7	41.5	45.2		29.6	33.3	45.8	44.4	33.0	29.4	22.4	29.5
Hydpho	**26.0**	**28.3**	**30.2**	**24.5**	**30.2**	**26.4**	**30.2**	**24.5**	**30.2**	**24.5**	**30.2**	**24.5**	**30.2**	**26.4**	**28.3**	**34.0**	**28.3**	**24.5**	**26.4**		**35.2**	**33.3**	**27.1**	**26.7**	**24.3**	**34.4**	**40.8**	**27.9**
Ala	7.4	9.4	7.6	3.8	9.4	5.7	9.4	3.8	9.4	3.8	9.4	3.8	9.4	5.7	7.6	5.7	5.7	3.8	5.7		5.6	14.8	8.3	3.2	5.8	7.7	15.6	4.3
Arg	**1.9**	**1.9**	**3.8**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**1.9**	**3.8**	**3.8**	**1.9**		**0.0**	**5.6**	**2.1**	**6.4**	**2.4**	**3.2**	**1.4**	**8.1**
Asn	11.1	9.4	9.4	7.6	9.4	7.6	9.4	7.6	9.4	7.6	9.4	7.6	9.4	7.6	11.3	13.2	7.6	11.3	11.3		1.9	3.7	4.2	8.0	9.3	4.1	2.0	7.4
Asp	**13.0**	**3.8**	**5.7**	**11.3**	**9.4**	**9.4**	**9.4**	**11.3**	**9.4**	**11.3**	**9.4**	**11.3**	**9.4**	**9.4**	**9.4**	**3.8**	**5.7**	**5.7**	**3.8**		**7.4**	**7.4**	**4.2**	**8.0**	**10.5**	**7.7**	**2.0**	**7.8**
Cys	11.1	11.3	11.3	11.3	11.3	11.3	11.3	11.3	11.3	11.3	11.3	11.3	11.3	11.3	11.3	11.3	11.3	11.3	11.3		11.1	11.1	16.7	9.5	7.8	5.4	6.1	2.7
Gln	**3.7**	**3.8**	**1.9**	**3.8**	**1.9**	**3.8**	**3.8**	**3.8**	**3.8**	**3.8**	**3.8**	**3.8**	**3.8**	**3.8**	**3.8**	**5.7**	**5.7**	**3.8**	**7.6**		**5.6**	**9.3**	**2.1**	**4.8**	**2.6**	**7.2**	**2.0**	**5.0**
Glu	5.6	9.4	5.7	7.6	5.7	7.6	5.7	5.7	5.7	7.6	5.7	5.7	5.7	7.6	5.7	5.7	5.7	3.8	3.8		5.6	3.7	2.1	4.8	5.8	7.7	0.7	9.3
Gly	**1.9**	**5.7**	**5.7**	**5.7**	**5.7**	**7.6**	**5.7**	**7.6**	**5.7**	**5.7**	**5.7**	**7.6**	**3.8**	**7.6**	**5.7**	**3.8**	**7.6**	**11.3**	**3.8**		**7.4**	**5.6**	**8.3**	**3.2**	**7.6**	**5.4**	**21.1**	**2.7**
His	1.9	3.8	3.8	3.8	1.9	3.8	1.9	3.8	1.9	3.8	1.9	3.8	1.9	3.8	1.9	1.9	5.7	1.9	1.9		5.6	1.9	0.0	1.6	1.6	2.3	0.0	1.6
Ile	**1.9**	**1.9**	**3.8**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**3.8**	**3.8**	**1.9**	**1.9**		**3.7**	**0.0**	**0.0**	**4.8**	**3.2**	**4.5**	**2.7**	**5.0**
Leu	3.7	5.7	7.6	7.6	7.6	7.6	7.6	7.6	7.6	7.6	7.6	7.6	7.6	7.6	7.6	5.7	3.8	5.7	3.8		9.3	1.9	4.2	3.2	4.1	7.7	3.4	5.4
Lys	**3.7**	**3.8**	**1.9**	**3.8**	**7.6**	**3.8**	**5.7**	**3.8**	**5.7**	**3.8**	**5.7**	**3.8**	**5.7**	**3.8**	**5.7**	**1.9**	**0.0**	**0.0**	**3.8**		**1.9**	**5.6**	**4.2**	**0.0**	**2.3**	**3.6**	**2.0**	**5.8**
Met	0.0	0.0	0.0	0.0	1.9	0.0	1.9	0.0	1.9	0.0	1.9	0.0	1.9	0.0	0.0	0.0	0.0	0.0	0.0		1.9	0.0	0.0	0.0	0.4	1.4	0.7	4.3
Phe	**5.6**	**3.8**	**1.9**	**1.9**	**3.8**	**1.9**	**3.8**	**1.9**	**3.8**	**1.9**	**3.8**	**1.9**	**3.8**	**1.9**	**3.8**	**3.8**	**5.7**	**0.0**	**3.8**		**11.1**	**7.4**	**0.0**	**4.8**	**1.6**	**6.8**	**0.0**	**3.9**
Pro	7.4	5.7	7.6	5.7	9.4	5.7	9.4	5.7	9.4	5.7	9.4	5.7	9.4	5.7	9.4	7.6	5.7	7.6	9.4		5.6	3.7	6.3	4.8	12.1	5.0	8.8	3.1
Ser	**3.7**	**3.8**	**5.7**	**5.7**	**3.8**	**3.8**	**3.8**	**5.7**	**3.8**	**5.7**	**3.8**	**5.7**	**5.7**	**3.8**	**3.8**	**3.8**	**3.8**	**7.6**	**9.4**		**1.9**	**3.7**	**10.4**	**6.4**	**6.7**	**4.5**	**6.8**	**7.0**
Thr	1.9	3.8	1.9	1.9	0.0	1.9	0.0	1.9	0.0	1.9	0.0	1.9	0.0	1.9	1.9	1.9	3.8	0.0	1.9		7.4	1.9	10.4	9.5	3.9	5.9	4.1	5.4
Trp	**1.9**	**1.9**	**1.9**	**1.9**	**1.9**	**1.9**	**1.9**	**1.9**	**1.9**	**1.9**	**1.9**	**1.9**	**1.9**	**1.9**	**7.6**	**1.9**	**1.9**	**1.9**	**1.9**		**3.7**	**1.9**	**0.0**	**1.6**	**1.1**	**1.4**	**6.8**	**0.4**
Tyr	7.4	5.7	5.7	7.6	1.9	7.6	1.9	7.6	1.9	7.6	1.9	7.6	1.9	7.6	1.9	3.8	5.7	7.6	3.8		1.9	3.7	2.1	6.4	2.7	2.3	1.4	2.0
Val	**5.6**	**5.7**	**7.6**	**9.4**	**7.6**	**9.4**	**7.6**	**9.4**	**7.6**	**9.4**	**7.6**	**9.4**	**7.6**	**9.4**	**7.6**	**13.2**	**7.6**	**11.3**	**9.4**		**1.9**	**7.4**	**14.6**	**9.5**	**8.3**	**6.3**	**12.2**	**8.9**
Σ Res.	54	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53		54	54	48	63	1917	221	147	258

Values reported as percentage of the number of amino acids within the indicated region.

Appendix H. Amino acid composition of *M. configurata* proteins with peritrophin A domains (McPPADs).

Protein	Domain	Repeat sequence	Number present
McIIM1	MD1	TAAP	22
	MD2	TAAP	1
		TPVAVVT	2
		TSAPIS	2
	MD3	PTVAPETTTA	3
		VTNA	4
McIIM2	MD1	TTEPTT	2
		PTTT	2
		TTTTTTPAPTT	8
McIIM3	MD1-2	TYTC	2
McIIM4	MD1	none	
	MD2	STTPV	12
		YEITTTPVSD	6
PxIIM	MD1	PEV(V/T)T	10
		TTL(A/E)PETDA	6
		VVTAIP	3
	MD2	PTTTESEAETVEVVTAIPTTTESEAETVEVVTAIPTTNAPETEATTVAETETPEVVTAV	3
	MD3	TVTAT(E/P/Q)EP(I/V)V	15
TnIIM22	MD1	TTQAP(A/T)	15
		AATTP	6
	MD2	TAAP	25

Appendix I. Tandem repeats within mucin domains of lepidopteran insect intestinal mucins.

Appendix I. (continued)

Protein	Domain	Repeat sequence	Number present
HaIIM2	MD1	TAAP	16
	MD2	TAAP	6
	MD3	TAAP	6
	MD4	TAAP	6
	MD5	TAAP	5
	MD6	AP(T/S)(T/S)TTTVAPT	6
HaIIM4	MD1	PEPTT	7
		TTTTTTTTTPAP	3
HaIIM5	MD1	PE(L/Q/P/A)TTTT	8
HaIIM86	MD1	TAAP(S/T)T	3
		TTTTTT(P/Q)(E/A)P	11
EpIIM1	MD1	PE(G/D)(S/N)GS(V/A)E	4
EpIIM2	MD1	TEAP(E/A)PE	14
EpIIM3	MD1	SGPE	2
BmIIM1	MD1	TQ(A/P)P	6
		TEQP	3
		T(A/E)PP	4
		T(A/V)AP	4
BmIIM2	MD1	(Q/L)APTTTT	4
		PTT	4
SeIIM	MD1	QAPTTTT	6